Routledge Encyclopedia of Philosophy

Philosophy of Language

Table of Contents

Adverbs

Adverbs are so named from their role in modifying verbs and other non-nominal expressions. For example, in 'John ran slowly', the adverb 'slowly' modifies 'ran' by characterizing the manner of John's running. The debate on the semantic contribution of adverbs centres on two approaches. On the first approach, adverbs are understood as predicate operators: for example, in 'John ran slowly', 'ran' would be taken to be a predicate and 'slowly' an operator affecting its meaning. Working this out in detail requires the resources of higher-order logic. On the second approach, adverbs are understood as predicates of 'objects' such as events and states, reference to which is revealed in logical form. For example, 'John ran slowly' would be construed along the lines of 'there was a running by John and it was slow', in which the adverb 'slowly' has become a predicate 'slow' applied to the event that was John's running.

Since adverbs are exclusively modifiers, they are classed among the syncategorematic words of terminist logic, the investigation of which carried the subject forward from Aristotle in the thirteenth century. (The contrasting 'categoremata' - grammatical subjects and predicates - are those words which have meaning independently.) They are of contemporary interest for philosophical logic and semantic theory, because particular accounts of them carry implications for the nature of combinatorial semantics and language understanding, and for ontology.

1 Syntactic types and semantic combination

There are several types of adverbial constructions, of which we distinguish the following classes: (a) 'manner' adverbs, which intuitively function as simple modifiers of verbs; (b) 'thematic' adverbs, of which some and possibly all function as (at least) two-place predicates in their own right; (c) adverbs of quantification, which express generality applying to whole sentences; and (d) discourse particles, whose meaning evidently derives from their role in linking clauses or independent sentences. (These categories are not exhaustive.) 'Adverbs', especially manner 'adverbs', are not in fact confined to single words. The general category is therefore not that of adverbs, but of adverbial phrases or adverbials (for example, 'more quickly than Mary', 'very frequently').

Typical manner adverbials are as in (1) below, thematic adverbs as in (2) and adverbs of quantification as in (3):

(1) John walked slowly/quietly/more quickly than Mary.
(2) Mary apparently/reluctantly went to New York.
(3) Mary occasionally/always walks to work.

Discourse particles, considered briefly below, include 'but', 'anyway' and several others. We discuss these cases in turn.

The essential logical problem of manner adverbials is already apparent in the simplest examples. A verb combines with a manner adverb to form a complex verbal construction of the same type. Thus 'walk' and 'walk slowly' are both predicates, and the syntax of the combination may be depicted as follows:

$[_V [_V \text{walk}] [_{Adv} \text{slowly}]]$

If (disregarding tense) we take 'walk' as a one-place predicate, then the semantics of this combination might be given by positing that 'slowly' is interpreted as a predicate operator; that is, as a function that maps one-place predicate interpretations onto other one-place predicate interpretations. Alternatively, it may be suggested that 'slowly' and the other manner adverbials are, logically speaking, predicates in their own right, specifically predicates of actions. The adjectives to which they are related do seem to play this role. Corresponding to (1), for instance, we have the adjectival predications

(4) John's walk was slow/quiet/quicker than Mary's.

If we take the further step of supposing that the verb 'walk' is in fact a two-place predicate, with a position for actions, then the combination 'walk slowly' can be interpreted as

walk(x, e) & slow(e),

where *e* ranges over actions. Comparing this account with the first alternative,

(slowly(walk))(*x*),

we see a trade-off: where predicates are taken to have a simple structure, the adverbial must be understood as an operator; but where extra structure, in the form of a place for actions, is posited, the semantic combination is truth-functional and predicate operators are not required.

The alternatives just sketched each have their defenders in the literature on adverbials. Adverbs are construed as predicate operators in formal theories of linguistic structure, including those of Montague (1974) and Lewis (1975). The predicative alternative was first advanced at length by Davidson (1967), and is elaborated by Parsons (1990). Semantic and metaphysical issues arise for each account; we take up some of these below.

Thematic adverbs are intuitively distinguished from manner adverbs in as much as they yield constructions adverb + verb which cannot be treated simply as new verbs: apparently going to New York is not a way of going to New York; and reluctantly going to New York is not a manner of travel, but an instance of travel whose agent was reluctant to undertake it. For the examples in (2) the following paraphrases suggest themselves:

It was apparent that Mary went to New York.
Mary went to New York and she was reluctant to go to New York.

The discourse particles, traditionally and appropriately called adverbs, have come under relatively formal study only in recent years. The following examples are representative:

(5) He was poor *but* honest.
(6) *Anyway*, I'm going to New York.

They resist analysis in terms of their contribution to truth-conditions, but carry implications for the discourses in which they occur, with (5) involving some presumptive contrast with what might have been expected and (6), as an assertion, functioning to indicate a return to a superordinate topic of conversation. See Levinson (1983) for examples and discussion.

2 Ambiguities

Many adverbs are ambiguous between thematic-adverbial and manner-adverbial interpretation. In an example such as

(7) Mary quickly objected,

we may have either the interpretation 'Mary's objection was delivered in a quick manner' or the interpretation 'Mary's objection came a short time after the enunciation of the proposition to which she objected'. The second, thematic-adverbial interpretation shows up in the corresponding adjective 'quick' in a construction such as

Mary was quick to object.

These examples suggest that grammatical appearance belies logical structure, since the thematic adverbial in (7) functions, logically speaking, as the main predicate of that sentence. Austin (1956) observed that adverbial position often disambiguates, with post-verb adverbs favouring the manner-interpretation, and pre-verb adverbs the thematic, as in the pair

He trod on the snail clumsily.
Clumsily, he trod on the snail.

Adverbs of quantification pick up arguments including, but not restricted to, the temporal. Sentences such as (8) are ambiguous, depending upon whether the quantification is over occasions, or over the subject:

(8) Travel books are seldom worth reading.

Where the adverb quantifies over time, (8) means that the occasions are few when it is worth reading travel books. But there is another salient interpretation:

Few travel books are worth reading.

Where quantification over time would be ridiculous in view of the subject matter, construal of the adverb with the subject is particularly salient, for example:

Quadratic equations seldom have real solutions.

The above reflections have it that thematic adverbs and adverbs of quantification are not modifiers at all, except in a purely grammatical sense. Inversely, there have been suggestions that what appear in language as if they were arguments to a predicate actually function as adverbial modifiers. Perhaps the best known account of this type is Roderick Chisholm's discussion of a certain class of statements about appearances. Chisholm suggests that a man who 'sees spots before his eyes' should be thought of as 'sensing in a spotty manner' or as someone whom things 'appeared to spottily' (1957). This philosophical move is designed to rid the locution of any implication that in sensing spots before his eyes the man is sensing a mysterious object, an appearance, which is before him, his eyes or his mind (see Mental states, adverbial theory of). In a similar vein, Goodman ([1968] 1976) construed the locution 'x represents an F', under the condition where there is no implication that such an F exists, as 'is an F-representation', effectively treating the predicate F adverbially.

The suggestions of Chisholm and Goodman, left as they are, become problematic if the project of giving a combinatorial semantics for language is taken seriously (see Compositionality). Suppose x is a picture and (9) is true, with no existential implications:

(9) x represents two unicorns galloping in a field.

If we write, with Goodman,

x is a two-unicorns-galloping-in-a-field representation,

then we have yet to provide a semantic structure to go along with the syntax. But there must be some such structure, since, for example, (9) obviously implies

x represents more than one unicorn in motion.

Similar issues arise for the predicate-operator theory of manner adverbs. For instance, that theory does not immediately deliver the obvious implication

(10) John walked slowly; therefore, John walked.

In Montague (1974) and much subsequent literature, this and similar implications are the consequences of semantic postulates in the sense of Carnap (1956). Even with such postulates, the relation between the adverbial constructions in (1) and their adjectival paraphrases (4) requires clarification. By contrast, the predicative theory proposed by Davidson is specifically designed to capture such implications and paraphrases. The premise of (10) is rendered as

(11) $(\exists e)(\text{walk}(\text{John}, e) \& \text{slow}(e))$,

that is, 'There was a walk by John and it was slow'; and the conclusion has the form

(12) $(\exists e)\text{walk}(\text{John}, e)$,

'There was a walk by John', a trivial implicate of the premise. For the paraphrase relation between (1) and (4), we have only to note that the complex noun 'John's walk' would be understood as a definite description of an action; that is, as '(the e) walk(John, e)'. 'John's walk was slow' then becomes

(13) slow((the e) walk(John, e)).

Given any standard treatment of the definite description, (13) will imply (11). However, the Davidsonian view is committed to supplying extra, hidden structure in all cases, and to taking events as individuals, a step that has often been considered metaphysically dubious.

Ensembles of adverbs show ambiguity of scope (see Scope). For example, one interpretation of (14) has John being clever in that he made a stupid response:

(14) John cleverly responded stupidly.

Modal adverbs such as 'necessarily' and 'contingently', which modify whole sentences, allow singular terms and quantifiers within their superficial scope to be interpreted as outside it (a point known to the terminist logicians). Tracking the relative scopes of modalities is part of the contemporary application of modal logic. As in the cases of manner and thematic adverbs, there are both operator-theories and predicate-theories of these expressions; that is, the modalities may be developed as one-place modal operators, with the same syntax as negation; or as predicates of sentences in a first-order formulation of the logical syntax of natural language (see Modal operators). Montague (1963) argued that the latter was unacceptable since the normal laws of modal logic could not all be maintained, on pain of paradox. The argument has subsequently been developed and discussed by others: see Koons (1992) for a survey and response.

3 Extensionality

Adverbial constructions of both the manner and thematic types, together with others, show a kind of superficial but persistent non-extensionality: superficial, because it may disappear under analysis; and persistent, because it may show up in places that the analysis itself uncovers. Predicate-operator theories can accept non-extensionality as the norm. Supposing, for example, that those who breathe are exactly those who perceive, it is absurd to infer that x perceives rapidly from the premise that x breathes rapidly. The predicate-operator theory, taking all operations to be in intension, has no such consequence. Or supposing that those who went to New York were exactly those who visited Times Square, it does not follow that if Mary reluctantly went to New York she also reluctantly visited Times Square.

For theories of Davidson's type, the constructions

x breathed rapidly
x perceived rapidly

are of the same logical type, but the events on which the adverb is predicated are different. The coextensiveness of 'breathe' and 'perceive' amounts to the coextensiveness of

(for some e) breathe(x, e)
(for some e) perceive(x, e).

Such coextensiveness no more implies the equivalence of 'x breathed rapidly' to 'x perceived rapidly' than the coextensiveness of 'x kicked something' and 'x saw something' would imply the equivalence of 'x saw something red' and 'x kicked something red'.

Non-extensionality is more troublesome when one considers events related as genus and species. For example, 'Mary flew slowly across Spain' does not imply 'Mary travelled slowly across Spain', even though any event of flying is an event of travelling. It follows that it is inadequate to represent 'Mary flew slowly across Spain' simply as

(for some e) (fly(Mary, e) & across Spain(e) & slow(e)).

Rather, we must add that the event e, which was a flying and therefore a travelling, was slow *for a flying*. These examples show at least that manner adverbials are relative to the sets of events against which a given event is evaluated. Other examples may show that it is not merely sets but also properties, not extensionally individuated, that form the background.

In any case, non-extensionality is evident in constructions with thematic adverbs. Davidson, following a suggestion by Hector-Neri Castañeda, considered examples such as that of

Oedipus intentionally married Jocasta,

from which it does not follow that Oedipus intentionally married his mother. In this case, the locus of non-extensionality, although due to the presence of the adverb 'intentionally', is the predicate 'intend', from which the adverb is derived. In this sense, it is independent of any peculiarities of adverbs, as in the Davidsonian paraphrase

Oedipus married Jocasta, and he intended to marry Jocasta.

Higginbotham (1989) explores a number of similar examples.

See also: Logical and mathematical terms, glossary of; Syntax §6

JAMES HIGGINBOTHAM

References and further reading

Austin, J.L. (1956) 'A Plea for Excuses', *Proceedings of the Aristotelian Society*; repr. in *Philosophical Papers*, Oxford: Clarendon Press, 1961, 123-52.(A famous and highly readable paper in which Austin tries to clarify the nature of action from the way in which we use excuses to escape responsibility for the consequences of what we do.)

Carnap, R. (1956) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Develops the idea that many apparently logical inferences are to be accounted for by semantic postulates that relate the meanings of the words in premises and conclusion; quite technical.)

Chisholm, R. (1957) *Perceiving: A Philosophical Study*, Ithaca, NY: Cornell University Press, esp. ch. 8. (Proposes an adverbial analysis of perception statements to avoid commitment to such entities as 'sense-data'.)

Davidson, D. (1967) 'The Logical Form of Action Sentences', in N. Rescher (ed.) *The Logic of Decision and Action*, Pittsburgh, PA: University of Pittsburgh Press; repr. in *Essays on Actions and Events*, Oxford: Clarendon Press, 1980, 105-48.(An analysis of adverbial modification as predication of events; readable and influential.)

Goodman, N. (1968) *Languages of Art*, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1976.(Proposes an adverbial analysis of the likes of 'is a picture of an *F*' to prevent the existence of a picture of an *F* entailing the existence of a pictured *F*.)

Higginbotham, J. (1989) 'Elucidations of Meaning', *Linguistics and Philosophy* 12: 465-517. (§2 defends a Davidsonian analysis in the context of a discussion of how lexical analysis interacts with permissible modes of syntactic composition; quite technical.)

Koons, R.C. (1992) *Paradoxes of Belief and Strategic Rationality*, Cambridge: Cambridge University Press. (Mainly concerns the role of common knowledge or mutual belief in generating liar-like paradoxes about rational action; very technical.)

Kretzmann, N. (1968) *William of Sherwood's Treatise on Syncategorematic Words*, Minneapolis, MN: University of Minnesota Press.(Elucidates the categorematic/syncategorematic distinction mentioned in the introduction to this entry.)

Levinson, S.C. (1983) *Pragmatics*, Cambridge: Cambridge University Press. (Includes discussion of words such as 'but' and 'anyway' which are traditionally regarded as adverbs but whose contribution to meaning is primarily at the level of pragmatics.)

Lewis, D.K. (1975) 'Adverbs of Quantification', in E. Keenan (ed.) *Formal Semantics of Natural Language*, Cambridge: Cambridge University Press, 3-15. (Semantics for adverbs construed as predicate operators.)

Montague, R. (1963) 'Syntactical Treatment of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability', *Acta Philosophica Fennica* 16; repr. in *Formal Philosophy*, ed. R. Thomason, New Haven, CT: Yale University Press, 1974, 286-302.(Shows that straightforward analysis of modal adverbs ('necessarily', 'possibly') as predicates of sentences ('is necessarily true', 'is possibly true') leads to the generation of a paradox similar to Tarski's in his proof of the 'undefinability of truth'; very technical.)

Montague, R. (1974) *Formal Philosophy*, ed. R. Thomason, New Haven, CT: Yale University Press.(Montague's collected papers in natural language semantics; extremely technical. For adverbs, see especially 'English as a Formal Language'.)

Parsons, T. (1990) *Events in the Semantics of English*, Cambridge, MA: MIT Press. (Extended development and discussion of Davidson's analysis of adverbs; clear and, for the most part, non-technical; good bibliography.)

Ambiguity

A word, phrase or sentence is ambiguous if it has more than one meaning. The word 'light', for example, can mean not very heavy or not very dark. Words like 'light', 'note', 'bear' and 'over' are lexically ambiguous. They induce ambiguity in phrases or sentences in which they occur, such as 'light suit' and 'The duchess can't bear children'. However, phrases and sentences can be ambiguous even if none of their constituents is. The phrase 'porcelain egg container' is structurally ambiguous, as is the sentence 'The police shot the rioters with guns'. Ambiguity can have both a lexical and a structural basis, as with sentences like 'I left her behind for you' and 'He saw her duck'.

The notion of ambiguity has philosophical applications. For example, identifying an ambiguity can aid in solving a philosophical problem. Suppose one wonders how two people can have the same idea, say of a unicorn. This can seem puzzling until one distinguishes 'idea' in the sense of a particular psychological occurrence, a mental representation, from 'idea' in the sense of an abstract, shareable concept. On the other hand, gratuitous claims of ambiguity can make for overly simple solutions. Accordingly, the question arises of how genuine ambiguities can be distinguished from spurious ones. Part of the answer consists in identifying phenomena with which ambiguity may be confused, such as vagueness, unclarity, inexplicitness and indexicality.

1 Types of ambiguity

Ambiguity is a property of linguistic expressions. A word, phrase or sentence is ambiguous if it has more than one meaning. Obviously this definition does not say what meanings are or what it is for an expression to have one (or more than one). For a particular language, this information is provided by a grammar, which systematically pairs forms with meanings, ambiguous forms with more than one meaning (see Semantics).

There are two types of ambiguity, *lexical* and *structural*. Lexical ambiguity is by far the more common. Everyday examples include nouns like 'chip', 'pen' and 'suit', verbs like 'call', 'draw' and 'run' and adjectives like 'deep', 'dry' and 'hard'. There are various tests for lexical ambiguity. One test is having two unrelated antonyms, as with 'hard', which has both 'soft' and 'easy' as opposites. Another is the conjunction reduction test. Consider the sentence, 'The tailor pressed one suit in his shop and one in the municipal court'. Evidence that the word 'suit' (not to mention 'press') is ambiguous is provided by the anomaly of the 'crossed interpretation' of the sentence, on which 'suit' is used to refer to an article of clothing and 'one' to a legal action.

The above examples of ambiguity are each a case of one word with more than one meaning. However, it is not always clear when we have only one word. The verb 'desert' and the noun 'dessert', which sound the same but are spelled differently, count as distinct words (they are homonyms). So do the noun 'bear' and the verb 'bear', even though they not only sound the same but are spelled the same. These examples may be clear cases of homonymy, but what about the noun 'respect' and the verb 'respect' or the preposition 'over' and the adjective 'over'? Are the members of these pairs homonyms or different forms of the same word? There is no general consensus on how to draw the line between cases of one ambiguous word and cases of two homonymous words. Perhaps the difference is ultimately arbitrary.

Sometimes one meaning of a word is derived from another. For example, the cognitive sense of 'see' (to see that something is so) seems derived from its visual sense. The sense of 'weigh' in 'He weighed the package' is derived from its sense in 'The package weighed two pounds'. Similarly, the transitive senses of 'burn', 'fly' and 'walk' are derived from their intransitive senses. Now it could be argued that in each of these cases the derived sense does not really qualify as a second meaning of the word but is actually the result of a lexical operation on the underived sense. This argument is plausible to the extent that the phenomenon is systematic and general, rather than peculiar to particular words. Lexical semantics has the task of identifying and characterizing such systematic phenomena. It is also concerned to explain the rich and subtle semantic behaviour of common and highly flexible words like the verbs 'do' and 'put' and the prepositions 'at', 'in' and 'to'. Each of these words has uses which are so numerous yet so closely related that they are often described as 'polysemous' rather than ambiguous.

Structural ambiguity occurs when a phrase or sentence has more than one underlying structure, such as the phrases 'Tibetan history teacher', 'a student of high moral principles' and 'short men and women', and the sentences 'The

girl hit the boy with a book' and 'Visiting relatives can be boring'. These ambiguities are said to be structural because each such phrase can be represented in two structurally different ways, for example '[Tibetan history] teacher' and 'Tibetan [history teacher]'. Indeed, the existence of such ambiguities provides strong evidence for a level of underlying syntactic structure (see Syntax). Consider the structurally ambiguous sentence, 'The chicken is ready to eat', which could be used to describe either a hungry chicken or a cooked chicken. It is arguable that the operative reading depends on whether or not the implicit subject of the infinitive clause 'to eat' is tied anaphorically to the subject ('the chicken') of the main clause.

It is not always clear when we have a case of structural ambiguity. Consider the elliptical sentence, 'Perot knows a richer man than Trump'. It has two meanings, that Perot knows a man who is richer than Trump and that Perot knows a man who is richer than any man Trump knows, and is therefore ambiguous. But what about the sentence 'John loves his mother and so does Bill'? It can be used to say either that John loves John's mother and Bill loves Bill's mother or that John loves John's mother and Bill loves John's mother. But is it really ambiguous? One might argue that the clause 'so does Bill' is unambiguous and may be read unequivocally as saying in the context that Bill does the same thing that John does, and although there are two different possibilities for what counts as doing the same thing, these alternatives are not fixed semantically. Hence the ambiguity is merely apparent and better described as semantic underdetermination.

Although ambiguity is fundamentally a property of linguistic expressions, people are also said to be ambiguous on occasion in how they use language. This can occur if, even when their words are unambiguous, their words do not make what they mean uniquely determinable. Strictly speaking, however, ambiguity is a semantic phenomenon, involving linguistic meaning rather than speaker meaning (see Meaning and communication). Generally when one uses ambiguous words or sentences, one does not consciously entertain their unintended meanings, although there is psycholinguistic evidence that when one hears ambiguous words one momentarily accesses and then rules out their irrelevant senses. When people use ambiguous language, generally its ambiguity is not intended. Occasionally, however, ambiguity is deliberate, as with an utterance of 'I'd like to see more of you' when intended to be taken in more than one way in the very same context of utterance.

2 Ambiguity contrasted

It is a platitude that what your words convey 'depends on what you mean'. This suggests that one can mean different things by what one says, but it says nothing about the variety of ways in which this is possible. Semantic ambiguity is one such way, but there are others: homonymy (mentioned in §1), vagueness, relativity, indexicality, nonliterality, indirection and inexplicitness. All these other phenomena illustrate something distinct from multiplicity of linguistic meaning.

An expression is vague if it admits of borderline cases (see Vagueness). Terms like 'bald', 'heavy' and 'old' are obvious examples, and their vagueness is explained by the fact that they apply to items on fuzzy regions of a scale. Terms that express cluster concepts, like 'intelligent', 'athletic' and 'just', are vague because their instances are determined by the application of several criteria, no one of which is decisive.

Relativity is illustrated by the words 'heavy' and 'old' (these are vague as well). Heavy people are lighter than nonheavy elephants, and old cats can be younger than some young people. A different sort of relativity occurs with sentences like 'Jane is finished' and 'John will be late'. Obviously one cannot be finished or late *simpliciter* but only finished with something or late for something. This does not show that the words 'finished' and 'late' are ambiguous (if they were, they would be ambiguous in as many ways as there are things one can be finished with or things one can be late for), but only that such a sentence is semantically underdeterminate - it must be used to mean more than what the sentence means.

Indexical terms, like 'you', 'here' and 'tomorrow', have fixed meaning but variable reference. For example, the meaning of the word 'tomorrow' does not change from one day to the next, though of course its reference does (see Demonstratives and indexicals).

Nonliterality, indirection and inexplicitness are further ways in which what a speaker means is not uniquely determined by what their words mean (see Speech acts §4). They can give rise to unclarity in communication, as might happen with utterances of 'You're the icing on my cake', 'I wish you could sing longer and louder', and 'Nothing is on television tonight'. These are not cases of linguistic ambiguity but can be confused with it because

speakers are often said to be ambiguous.

3 Philosophical relevance

Philosophical distinctions can be obscured by unnoticed ambiguities. So it is important to identify terms that do double duty. For example, there is a kind of ambiguity, often described as the 'act/object' or the 'process/product' ambiguity, exhibited by everyday terms like 'building', 'shot' and 'writing'. Confusions in philosophy of language and mind can result from overlooking this ambiguity in terms like 'inference', 'statement' and 'thought'. Another common philosophical ambiguity is the type/token distinction. Everyday terms like 'animal', 'book' and 'car' apply both to types and to instances (tokens) of those types. The same is true of linguistic terms like 'sentence', 'word' and 'letter' and of philosophically important terms like 'concept', 'event' and 'mental state' (see Type/token distinction).

Although unnoticed ambiguities can create philosophical problems, ambiguity is philosophically important also because philosophers often make spurious claims of it. Indeed, the linguist Charles Ruhl (1989) has argued that certain ostensible ambiguities, including act/object and type/token, are really cases of lexical underdetermination. Kripke (1977) laments the common stratagem, which he calls 'the lazy man's approach in philosophy', of appealing to ambiguity to escape from a philosophical quandary, and Grice (1967) urges philosophers to hone 'Modified Occam's Razor: senses are not to be multiplied beyond necessity'. He illustrates its value by shaving a sense off the logical connective 'or', often thought to have both an inclusive and exclusive sense. Grice argues that, given its inclusive meaning, its exclusive use can be explained entirely on pragmatic grounds (see Implicature §6; Pragmatics §12). Another example, prominent in modern philosophy of language, is the ambiguity alleged to arise from the distinction between referential and attributive uses of definite descriptions (see Descriptions §5). Less prominent but not uncommon is the suggestion that pronouns are ambiguous as between their anaphoric and deictic use. So, for example, it is suggested that a sentence like 'Oedipus loves his mother' has two 'readings' - that is, it is ambiguous - because it can be used to mean either that Oedipus loves his own mother or that Oedipus loves the mother of some contextually specified male. However, this seems to be an insufficient basis for the claim of ambiguity. After all, being previously mentioned is just another way of being contextually specified. Accordingly, there is nothing semantically special in this example about the use of 'his' to refer to Oedipus.

Claims of structural ambiguity can also be controversial. Of particular importance are claims of scope ambiguity, which are commonly made but rarely defended (see Scope). A sentence like 'Everybody loves somebody' is said to exhibit a scope ambiguity because it can be used to mean either that for each person, there is somebody that that person loves or (however unlikely) that there is somebody that everybody loves. These uses may be represented, respectively, by the logical formulas ' $(\forall x)(\exists y)(Lxy)$ ' and ' $(\exists y)(\forall x)(Lxy)$ '. It is generally assumed that, because different logical formulas are needed to represent the different ways in which an utterance of such a sentence can be taken, the sentence itself has two distinct logical forms (see Logical form). Sustaining this claim of ambiguity requires identifying a level of linguistic description at which the sentence can be assigned two distinct structures. Some grammarians have posited a level of LF, corresponding to what philosophers call logical form, at which relative scope of quantified noun phrases may be represented. However, LF of this kind does not explain scope ambiguities that philosophers attribute to sentences containing modal operators and psychological verbs, such as 'The next president might be a woman' and 'Ralph wants a sloop'. An utterance of such a sentence can be taken in either of two ways, but it is arguable that the sentence is not ambiguous but merely semantically underdeterminate with respect to its two alleged 'readings'.

Notwithstanding the frequency in philosophy of unwarranted and often arbitrary claims of ambiguity, it cannot be denied that some terms really are ambiguous. The nouns 'bank' and 'suit' are clear examples and so are the verbs 'bank' and 'file'. Philosophers sometimes lament the prevalence of ambiguity in natural languages and yearn for an ideal language in which it is absent. But ambiguity is a fact of linguistic life. Despite the potentially endless supply of words, many words do double duty or more. And despite the unlimited number of sentences, many have several meanings and their utterance must be disambiguated in the light of the speaker's likely intentions.

See also: Language, philosophy of; Semantics

KENT BACH

References and further reading

Atlas, J.D. (1989) *Philosophy Without Ambiguity: A Logico-Linguistic Essay*, Oxford: Oxford University Press. (Examines ambiguity tests and questions certain philosophical appeals to ambiguity.)

Bach, K. (1994) 'Conversational Impliciture', *Mind and Language* 9: 124-62. (Identifies ways, distinct from those identified by Grice, in which linguistic meaning can underdetermine speaker meaning and, in particular, contrasts semantic underdetermination with ambiguity.)

Cruse, D.A. (1986) *Lexical Semantics*, Cambridge: Cambridge University Press. (Chapter 3 discusses linguistic features of ambiguity and examines tests for it.)

Grice, H.P. (1967) 'Logic and Conversation', *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 1989.(Chapter 2 of this collection explains how multiple uses of an expression can often be explained without appealing to ambiguity.)

Kripke, S.A. (1977) 'Speaker's Reference and Semantic Reference', *Midwest Studies in Philosophy* 2: 255-76.(A case study, focusing on the distinction between referential and attributive uses of definite descriptions, in how to expose a gratuitous philosophical claim of ambiguity.)

May, R. (1985) *Logical Form: Its Structure and Derivation*, Cambridge, MA: MIT Press.(Investigates the hypothesis within generative grammar that there is a level of representation, Logical Form.)

Ruhl, C. (1989) *On Monosemy: A Study in Linguistic Semantics*, Albany, NY: State University of New York Press.(Argues for the presumption that a word has a single, though highly abstract, meaning rather than the multiplicity of meanings commonly attributed, as by the list of definitions in its dictionary entry.)

Zwicky, A. and Sadock, J. (1975) 'Ambiguity Tests and How to Fail Them', in J. Kimball (ed.) *Syntax and Semantics*, vol. 4, New York: Academic Press.(Presents and assesses various linguistic tests of ambiguity.)

Analyticity

In Critique of Pure Reason Kant introduced the term 'analytic' for judgments whose truth is guaranteed by a certain relation of 'containment' between the constituent concepts, and 'synthetic' for judgments which are not like this. Closely related terms were found in earlier writings of Locke, Hume and Leibniz. In Kant's definition, an analytic judgment is one in which 'the predicate B belongs to the subject A, as something which is (covertly) contained in this concept A' ([1781/1787] 1965: 48). Kant called such judgments 'explicative', contrasting them with synthetic judgments which are 'ampliative'. A paradigmatic analyticity would be: bachelors are unmarried. Kant assumed that knowledge of analytic necessities has a uniquely transparent sort of explanation. In the succeeding two centuries the terms 'analytic' and 'synthetic' have been used in a variety of closely related but not strictly equivalent ways. In the early 1950s Morton White (1950) and W.V. Quine (1951) argued that the terms were fundamentally unclear and should be eschewed. Although a number of prominent philosophers have rejected their arguments, there prevails a scepticism about 'analytic' and the idea that there is an associated category of necessary truths having privileged epistemic status.

1 The attack on 'analytic'

'Analytic' has been used in a wide variety of ways: truth by conceptual containment and truth whose denial is contradictory (Kant 1781/1787); logical truth (Bolzano 1837; Feigl 1949); truth by definition and logical derivation (Frege 1884; Pap 1958); truth in virtue of form (Schlick 1930-1); truth by definition and logical truth (Carnap 1937, 1947); truth by definition (Ayer 1936); truth based on meaning (Ayer 1936; C.I. Lewis 1944); truth by semantical rule (Carnap 1947); truth in all possible worlds (C.I. Lewis 1944; D.K. Lewis 1969); convertibility into logical truth by substitution of synonyms (Quine 1951); truth by implicit convention (Putnam 1962); and so on. Although related, not all of these uses are equivalent. For example, logical truths are not true by definition (in the sense of explicit definition), but they are trivially true by definition *plus* logic. Furthermore, Gödel's incompleteness result shows that logical derivability and logical truth are not equivalent (see Gödel's theorems §3). Likewise various principles (for example, supervenience principles; see Supervenience) which are true in all possible worlds seem not to be true by definition plus logic (if 'definition' does not include 'implicit definitions'; see Definition). Similarly, it may be doubted that correct definitions provide exact synonyms (see §4). Little care has been taken to distinguish these disparate uses and needless confusions have resulted. But, as Strawson and Grice (1956) note, observations of this sort 'would scarcely amount to a rejection of the distinction [as Quine urges]. They would, rather, be a prelude to clarification'.

In 'Two Dogmas of Empiricism' (1951), Quine went far beyond the call for clarification; he argued that there simply was no distinction between analytic and synthetic truths. Quine's argument is an enthymeme with roughly the following form: there is no non-circular, purely *empiricist* clarification of the distinction, and therefore there is no distinction at all: 'That there is such a distinction to be drawn at all is an unempirical dogma of empiricists, a metaphysical article of faith' (see Quine, W.V. §8).

2 Extending the attack to 'necessary' and 'a priori'

Quine's attack did not stop with 'analytic'. Following the lead of logical positivists, he (wrongly) held that, if there were any such distinctions, the analytic/synthetic distinction would be the same as the necessary/contingent distinction, which in turn would be the same as the a priori/a posteriori distinction. So, on his view, since there is no analytic/synthetic distinction, there is no necessary/contingent distinction and no a priori/a posteriori distinction either. It is an error, however, to equate these distinctions, so these further conclusions do not follow.

Quine's attack on the necessary/contingent distinction has not convinced many. Most philosophers accept the distinction and invoke it in their work; modal logic and modal metaphysics are thriving subjects (see Modal logic; Modal logic, philosophical issues in). Do modal notions have non-circular definitions? Although definitions have been suggested, such as Church (1951) and Bealer (1982), there would be nothing unreasonable in holding them to be primitive (Prior and Fine 1977). After all, everyone must take *some* notions to be primitive.

What *epistemic justification* do we have for the necessary/contingent distinction? No doubt Quine was right that it cannot be justified on purely empirical grounds (that is, using only phenomenal experience and/or observation).

But justification is not always purely empirical: logicians, mathematicians, linguists and philosophers rely heavily on *intuitions* in the justification of their theories. (Quine himself relies on intuition in defending his formulations of set theory over others.) And there are no convincing arguments that such use of intuitions is illegitimate. Once one acknowledges intuitive evidence, the necessary/contingent distinction has a straightforward justification: we have a very wide range of robust modal intuitions (for example, the intuition that it is contingent that the number of planets is greater than seven - there could have been fewer), and when such intuitions are taken as evidence, the best theory is one which accepts the distinction at face value.

The necessary/contingent distinction is a metaphysical distinction. The a priori/a posteriori distinction, by contrast, is epistemological. Indeed, Kripke and Putnam have convincingly argued that there are necessary truths (Hesperus is Phosphorus, water is H_2O, and so on) which are impossible to justify a priori (see Kripke, S.A. §3; Putnam, H. §3). While the former distinction might well lack a non-circular analysis, the latter certainly has one: *p* is a piece of a posteriori evidence if and only if *p* is the content of an experience (a phenomenal experience, observation, memory or testimony); *p* is a piece of a priori evidence if and only if *p* is a piece of evidence which is not the content of an experience in the relevant sense. (On the assumption that one's intuitions are evidence, the contents of those intuitions would constitute one's a priori evidence.) A theory has an a posteriori justification if and only if the evidence on which that justification is based is a posteriori; a theory has an a priori justification if and only if the evidence on which that justification is based is a priori. Note that this sort of analysis does not presuppose that any theory has an a priori justification; one could deny that there is any a priori evidence. The point is that, contrary to Quine's allegation, the *notion* of a priori justification has a straightforward analysis (see A posteriori; A priori).

Quine seems to assume that a priori judgments would need to be infallible and unrevisable. But this is mistaken, as is evident from our ongoing a priori theorizing about the logical paradoxes. One of the main traditional lines of thought on the a priori - from Plato to Gödel - recognizes that a priori justification is fallible and holistic, relying on dialectic and/or a priori theory construction.

3 Defending 'analytic' against Quine

Unlike 'necessary', 'analytic' is a technical term. Accordingly, it is legitimate to demand, as Quine (1951) does, that its use be explained. This could be done with examples, but none of the (nonequivalent) notions listed above fits the standard examples perfectly, nor does any stand out as *the* most salient. This leaves the alternative of giving a definition. Quine's view is that *not one* of the historically prominent uses of 'analytic' (except the 'logical truth' use) has a satisfactory definition. But this radical scepticism is unwarranted.

Consider the following definition: a necessarily true sentence is analytic if and only if it may be converted into a logically true sentence by replacing its syntactically simple predicates with predicates which mean exactly the same thing. This seems to respect the idea that analytic truths are a priori and moreover are justifiable in a specially simple way. Quine (1960) would object to this definition by appeal to his thesis of the indeterminacy of translation - the thesis that 'there is no fact of the matter' concerning claims about identity of meaning (see Radical translation and radical interpretation §§1-3). His arguments for this thesis have not convinced many philosophers, however, for they depend on quasi-verificationist or behaviourist premises. Most philosophers reject Quine's scepticism about meaning, realism having become the dominant view. One reason for this is the advent of the broadly Gricean picture (Grice 1989) according to which meaning is analysable in terms of the propositional attitudes. Accordingly, if there is a fact of the matter about the latter, there is about the former (see Communication and intention). And, since the cognitivist revolution in psychology and philosophy of mind, nearly everyone is a realist about propositional attitudes. Thus, at least one use of 'analytic sentence' ought to be acceptable to these realist philosophers.

The same moral holds for at least certain uses of 'analytic' as it applies to propositions. Despite Quine's scepticism, most philosophers have become convinced that in logic, psychology and semantics there is need for structured propositions, that is, propositions which have a logical form (or sense structure). This makes possible a definition of 'analytic' in another one of its standard uses (Katz 1986; Bealer 1982): *p* is analytic iff$_{def}$ every proposition having the same form (structure) as *p* is necessary.

4 More damaging problems with 'analytic'

Although the above definition of 'analytic' is cogent, the term so-defined fails to apply to a number of examples which traditionally would have been deemed 'analytic'; the definition is too narrow. For example, the defined use does not cover Kant's paradigm example of an analyticity, namely, that bodies are extended. To accommodate this and a wide array of other examples (that circles are curves and so forth; see below), one must turn to a wider definition of 'analytic' - one relying on some philosophically robust notion such as definition, conceptual analysis, or the kind of meaning relations which hold between a definiendum and a definiens or between an analysandum and an analysans. For example, it is at least credible that there is a definition of 'body' in one of its senses according to which 'Bodies are extended' would be true by definition. Unfortunately, these wider accounts of 'analytic' give rise to a complementary problem: they let in too much.

The following familiar definitions illustrate the problem: x is a circle if and only if x is a closed plane figure every point on which is equidistant from a common point; x is a circle if and only if x is a closed plane figure every arc of which has equal curvature. There seems to be nothing to recommend one over the other; if either is a correct definition, both are. In that case it seems that the following would be true by definition plus logic: x is a closed plane figure every point on which is equidistant from a common point if and only if every arc of x has equal curvature. But Kant would deem this biconditional 'ampliative': in any standard axiomatic formulation of geometry, the proof of it would require axioms and axioms - as opposed to definitions - are supposed to be synthetic. Evidently, this argument can be adapted to many other a priori necessities traditionally thought to be paradigmatically synthetic.

A related kind of problem arises in connection with conceptual analysis. One of the most celebrated conceptual analyses in mathematical philosophy is the classical analysis of effective calculability, or computability. On Church's version, a function is effectively calculable if and only if it is lambda-calculable. On Turing's version, a function is effectively calculable if and only if it is Turing computable (see Church's thesis). Most philosophers deem each version to be a successful conceptual analysis. But when the two analyses are combined, it follows immediately that a function is recursive if and only if it is Turing computable. But this biconditional - which is an important 'ampliative' theorem of formal number theory - would then turn out to be analytic (in the sense of being true by conceptual analysis plus logic) even in the event that logicism is false.

These problems suggest that there is no coherent way to draw an analytic/synthetic distinction along the lines Kant thought. One response is to settle for a severely restricted use of 'analytic' which concerns only concepts with unique structures (that is, unique 'decompositions'). The price of this move, however, is high: the vast majority of our concepts - including nearly all of the concepts philosophers have sought to define or analyse (good, true, valid, number, meaning, knowledge, and so on) - are in this sense unstructured and so would not give rise to new analyticities. At best, rather uninteresting concepts (such as bachelor) are of this sort. In consequence, even if knowledge of analyticities (in the narrow sense) had a transparent epistemic explanation as Kant assumed, the sort of knowledge one seeks in typical philosophical definitions or analyses would need quite another sort of explanation.

5 Epistemological problems with 'analytic'

Kant and his successors simply assumed that knowledge of analyticities has a transparent sort of explanation (often linked to a simplistic 'pictorial' or 'mereological' view of concepts). Just what that explanation is supposed to be has never been satisfactorily stated. It cannot be that analytic propositions are those whose truth is recognized just by virtue of possessing the constituent concepts, for *no* proposition is like this. For example, it is in principle possible that someone who possesses the relevant concepts but who is in sufficiently defective cognitive conditions (deficient intelligence, attentiveness, and so on) might fail to recognize that all and only bachelors are unmarried men. It does no good to relax this account by holding that analytic propositions are those whose truth would be recognized by anyone in sufficiently good cognitive conditions just by virtue of possessing the constituent concepts, for this lets in too much: anyone in sufficiently good cognitive conditions could not fail to recognize, say, that figure A is congruent with figure B if and only if B is congruent with A. But such propositions are the very paradigms of what Kant would have deemed synthetic a priori and requiring a different sort of explanation. (For these same reasons, purely epistemic definitions of 'analytic' are problematic.)

Another line of explanation is to liken our knowledge, say, that bachelors are men to our knowledge that unmarried men are men, or more generally, that ABs are Bs - that is, to liken this knowledge to our knowledge of a certain kind of *logical* truth. But how do we know the logical truth that ABs are Bs? Is the explanation fundamentally different from the explanation of our knowledge of other kinds of logical truths - for example, that Bs are As or Bs? It is hard to see why it should be. This raises the question of how we know logical truths generally. For instance, is the explanation of our (logical) knowledge that identity is a symmetric relation ($A = B$ iff $B = A$) really different from the explanation of our (nonlogical) knowledge that congruence is a symmetric relation? From a phenomenological point of view, both instances of knowing (logical and nonlogical) arise from a priori intuitions and these intuitions, phenomenologically, are not relevantly different. On this score, therefore, there is no reason to think that our a priori knowledge divides neatly into two kinds having quite different explanations.

On the contrary, there is a promising unified explanation of a priori knowledge generally, which goes roughly as follows. In every case a priori knowledge is based evidentially on a priori intuition. The evidential force of a priori intuition is to be explained in terms of a general analysis of concept possession: it is constitutive of concept possession that in suitably good cognitive conditions intuitions regarding the behaviour of the concept need to be largely correct. If in suitably good cognitive conditions one did not have such intuitions, one would not be said to possess the concept. If something like this is right, then, although they mark cogent logical and metaphysical distinctions, all the listed uses of 'analytic' - even the narrow uses - fail to mark an epistemically significant category of knowledge.

The picture that results is complicated somewhat by Kripke and Putnam's doctrine that there are essentially a posteriori necessities - for example, water = H_2O. Among these are some which may plausibly be deemed *scientific definitions*. If 'analytic' is used in the sense of truth by definition plus logic, where 'definition' is understood to include scientific definitions, then there would be necessities which would be both analytic and essentially a posteriori. Evidently the Kripke-Putnam doctrine applies only to 'semantically unstable' expressions - that is, expressions ('water', 'gold', 'heat', and so on) whose meaning could be different in some population of speakers whose epistemic situation is qualitatively identical to ours. These are expressions to whose meaning the external environment makes some contribution (see Content: wide and narrow). The above picture, however, holds straightforwardly for semantically stable expressions ('conscious', 'know', 'good', and so on) which loom large in philosophical analysis.

6 Whither 'analytic'?

At this stage, one may reasonably ask whether continued use of 'analytic' serves any purpose in philosophy. Although the term evidently lacks the epistemological significance once attributed to it, the wider use of 'analytic' in the sense of true by definition plus logic still has utility, namely, in posing an important question: Are there necessary truths (supervenience principles, the incompatibility of colours, and so on) which are not analytic in this sense? The answer appears to be affirmative if 'definition' is understood straightforwardly as ordinary explicit definition. But if, as some have proposed, 'definition' is understood to include 'implicit definitions', the answer is controversial and depends on what information may legitimately be loaded into 'implicit definitions'. On pain of trivializing significant traditional questions, however, surely not everything is admissible. Plainly there still are unanswered questions here. But they are really about the nature of definitions; 'analytic' does no work.

See also: Carnap, R.; Concepts; Conceptual analysis; Intensional entities; Kant, I.; Logical positivism; Necessary truth and convention

GEORGE BEALER

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz.(Uses of 'analytic' mentioned in §1.)
Bealer, G. (1982) *Quality and Concept*, Oxford: Clarendon Press.(Definition of 'analytic' in terms of propositions and concepts discussed in §3.)
Bealer, G. (1996) 'A Priori Knowledge and the Scope of Philosophy', *Philosophical Studies* 81: 121-42.(Account of a priori justification and knowledge in terms of intuition and concept possession described in §§2 and 5.)
Boghossian, P. (1997, forthcoming) 'Analyticity', in C. Wright and B. Hale (eds) *A Companion to the Philosophy*

of Language, Oxford: Blackwell.(Epistemic account of analyticity.)

Bolzano, B. (1837) *Wissenschaftslehre*, Sulzbach: Seidel, 4 vols; trans. and ed. R. George, in *Theory of Science*, Oxford: Blackwell, 1972; trans. B. Terrell, ed. J. Berg, in *Theory of Science*, Dordrecht: Reidel, 1973.(Use of 'analytic' mentioned in §1. The George and Terrell translations provide selections from Bolzano's work.)

Carnap, R. (1937) *Logical Syntax of Language*, London: Routledge.(Uses of 'analytic' mentioned in §1.)

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Uses of 'analytic' mentioned in §1.)

Church, A. (1951) 'A Formulation of the Logic of Sense and Denotation', in P. Henle, H.H. Kallen and S.K. Langer (eds) *Structure, Method and Meaning: Essays in Honor of Henry M. Scheffer*, New York: Liberal Arts Press, 3-24.(Offers non-circular definitions of modal notions.)

Feigl, H. (1949) 'Logical Empiricism', in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 3-26.(Use of 'analytic' mentioned in §1.)

Fine, K. (1994) 'Essence and Modality', *Philosophical Perspectives* 8: 1-16.(Account of analyticity and definition in terms of the notion of essence.)

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980.(Use of 'analytic' mentioned in §1.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.(Analysis of meaning in terms of propositional attitudes discussed in §3.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, New York: St Martin's Press, 1965. (Provides the original definition of 'analytic' and 'synthetic' and an account of the possibility of synthetic a priori knowledge.)

Katz, J.J. (1986) *Cogitations*, Oxford: Oxford University Press.(Account of analyticity in terms of sense structure mentioned in §§3 and 5.)

Lewis, C.I. (1944) 'The Modes of Meaning', *Philosophy and Phenomenological Research* 4: 236-49; repr. L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University if Illinois Press, 1952, 50-63. (Uses of 'analytic' mentioned in §1.)

Lewis, D.K. (1969) *Convention*, Cambridge, MA: Harvard University Press.(Use of 'analytic' mentioned in §1.)

Pap, A. (1958) *Semantics and Necessary Truth*, New Haven, CT: Yale University Press.(Use of 'analytic' mentioned in §1 and extended discussion of analyticity and related issues.)

Parsons, C. (1997, forthcoming) *Quine and Gödel on Analyticity*, in P. Leonardi and M. Santambrogio (eds) *On Quine*, Cambridge: Cambridge University Press.(Discussion of Gödel's uses of 'analytic' in relation to Quine.)

Prior, A.N. and Fine, K. (1977) *Worlds, Times and Selves*, London: Duckworth. (Defends thesis that modal notions are primitive.)

Putnam, H. (1962) 'The Analytic and the Synthetic', in H. Feigl and G. Maxwell (eds) *Minnesota Studies in the Philosophy of Science*, vol. 3, Minneapolis, MN: University of Minnesota Press; repr. H. Putnam, *Mind, Language and Reality: Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press, 1975, 33-69. (Use of 'analytic' mentioned in §1.)

Quine, W.V. (1951) 'Two Dogmas of Empiricism', *Philosophical Review* 60: 20-43; repr. in *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 1953, 20-46. (Classic attack on the analytic/synthetic distinction.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Defence of indeterminacy of translation and attack on the ontology of propositions discussed in §3.)

Quine, W.V. (1970) *Philosophy of Logic*, Englewood Cliffs, NJ: Prentice Hall.(Attack on the ontology of propositions discussed in §3.)

Schlick, M. (1930-1) *Wissenschftlicher Jahresbericht der Philosophischen Gesellschaft an der Universität zu Wein für das Vereinsjahr 1930/1*, repr. as 'Is There a Factual A Priori?', in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, 1949, 277-94.(Use of 'analytic' mentioned in §1.)

Strawson, P.F. and Grice, H.P. (1956) 'In Defense of a Dogma', *Philosophical Review* 65:141-58; repr. in H.P. Grice, *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 1989, 196-212.(Early response to Quine mentioned in §1.)

White, M. (1950) *The Analytic and the Synthetic: An Untenable Dualism*, in S. Hook (ed.) *John Dewey: Philosopher of Science and Freedom*, New York: The Dial Press, 316-30; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University of Illinois Press, 1951, 272-86.(Classic attack on the analytic/synthetic distinction.)

Anaphora

Anaphora describes a dependence of the interpretation of one natural language expression on the interpretation of another natural language expression. For example, the pronoun 'her' in (1) below is anaphorically dependent for its interpretation on the interpretation of the noun phrase 'Sally' because 'her' refers to the same person 'Sally' refers to.

(1) Sally likes her car.

As (2) below illustrates, anaphoric dependencies also occur across sentences, making anaphora a 'discourse phenomenon':

(2) A farmer owned a donkey. He beat it.

The analysis of anaphoric dependence has been the focus of a great deal of study in linguistics and philosophy. Anaphoric dependencies are difficult to accommodate within the traditional conception of compositional semantics of Tarski and Montague precisely because the meaning of anaphoric elements is dependent on other elements of the discourse.

Many expressions can be used anaphorically. For instance, anaphoric dependencies hold between the expression 'one' and the indefinite noun phrase 'a labrador' in (3) below; between the verb phrase 'loves his mother' and a 'null' anaphor (or verbal auxiliary) in (4); between the prepositional phrase 'to Paris' and the lexical item 'there' in (5); and between a segment of text and the pronoun 'it' in (6).

(3) Susan has a labrador. I want one too.
(4) John loves his mother. Fred does too.
(5) I didn't go to Paris last year. I don't go there very often.
(6) One plaintiff was passed over for promotion. Another didn't get a pay increase for five years. A third received a lower wage than men doing the same work. But the jury didn't believe any of it.

Some philosophers and linguists have also argued that verb tenses generate anaphoric dependencies.

1 Syntactic and static semantic approaches

An expression is 'anaphoric' if it depends for its interpretation on the interpretation of another expression. For example, the pronoun 'her' in (1) below is anaphorically dependent on the noun phrase 'Sally'.

(1) Sally likes her car.

Anaphoric dependencies also occur across sentences, making anaphora a discourse phenomenon:

(2) A farmer owned a donkey. He beat it.

Though anaphoric dependencies are a semantic phenomenon, syntax has a very important role to play in their analysis. In the early sixties, syntacticians noticed that anaphoric dependencies were subject to peculiar constraints that syntax could explain. For instance, consider (3) and (3′):

(3) John hit himself.
(3′) John hit him.

In (3) 'John' and 'himself' must refer to the same thing (they must be 'coreferential'), while in (3′) 'John' and 'him' cannot do so. Such observations prompted Chomsky (1981) to devise a theory of binding for anaphoric expressions, whereby all noun phrases in the syntactic representation are assigned an index. Reflexive pronouns are subject to a constraint that forces 'himself' in (3) to be co-indexed with and so bound to some noun phrase (NP) that occurs within the same clause, while nonreflexive pronouns obey another constraint that implies that 'him' in (3′) cannot be bound by an NP within the same clause in which it occurred.

While Chomsky and other syntacticians recognized that there is a semantic component to anaphora, they never

clearly articulated the semantic interpretation of co-indexed and non-co-indexed NPs. Semantics furnishes a translation of a syntactic structure into a logical representation, a 'logical form', that can then be assigned a model-theoretic interpretation using the formal semantics for the logical representation language, typically the language of first-order or higher-order logic (see Model theory). The syntactically generated indices on NPs are interpreted as indices on variables. Pronouns introduce an indexed variable in the logical form, while noun phrases introduce a generalized quantifier, which in turn introduces a bound variable with the index of the NP. The quantifier binds all occurrences of the same variable in its scope. So a sentence such as (1) can be represented as follows:

(1′) $\exists x_j \exists x_i (x_i = \text{Sally \& car}(x_j) \,\&\, \text{owns}(x_i, x_j) \,\&\, \text{likes}(x_i, x_j))$.

This approach works well enough for the simple cases of intra-sentential anaphora. But anaphoric links between pronouns and their antecedents also occur across sentences. For example, in giving the logical form of (2) above, anaphoric pronouns cause a problem. Formally, the first sentence yields (ignoring tense):

(2a) $\exists x(\text{farmer}(x) \,\&\, \exists y(\text{donkey}(y) \,\&\, \text{owns}(x, y)))$.

We now tackle the second sentence. The pronouns in the second sentence are anaphorically bound to the noun phrases in the first sentence, and so we should translate the pronouns as variables that are to be bound by the quantifiers introduced by the noun phrases in the first sentence. But we have already finished the translation of the first sentence and so closed off the rightward scope of the quantifiers. Conjoining the translation of the second sentence does not produce a bound variable reading of the variables introduced by the pronouns. It produces an open sentence:

(2′) $\exists x(\text{farmer}(x) \,\&\, \exists y(\text{donkey}(y) \,\&\, \text{owns}(x, y))) \,\&\, \text{beats}(x, y)$.

Montague Grammar (see Montague, R.M.) seems to solve this problem with the procedure of 'quantifying in' (Gamut 1991). We first postulate a syntactic rule on which the noun phrase is replaced in the syntactic tree by a pronoun with an appropriate index, and the original noun phrase is adjoined with the sentence. We then apply the Montague Grammar translation procedure. Now if we interpret a sequence of two sentences as a conjunction, we can then use the quantifying-in procedure to interpret (2). First, we quantify in the two noun phrases 'a farmer' and 'a donkey'. The syntactic rule for quantifying in produces the following structure for (2):
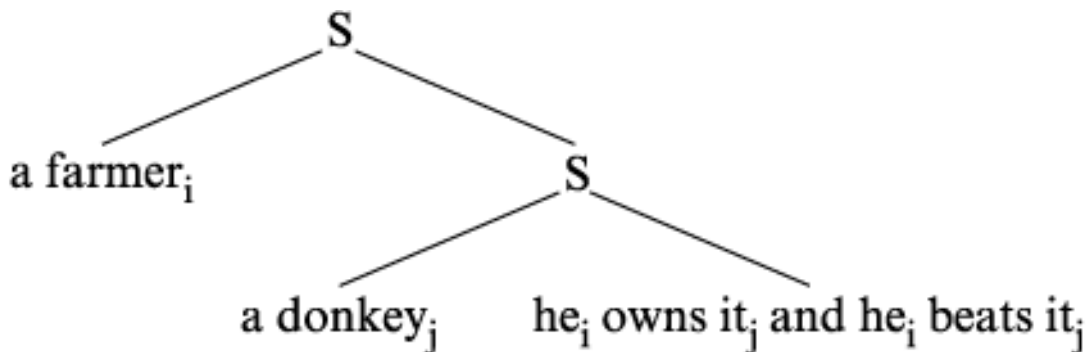
This tree now yields in Montague Grammar the correct logical form for (2):

(2*) $\exists x_j \exists x_i (\text{farmer}(x_i) \,\&\, \text{donkey}(x_j) \,\&\, \text{owns}(x_i, x_j) \,\&\, \text{beats}(x_i, x_j))$.

However, if we wish to continue to use the procedure of quantifying in to deal with discourses in which anaphoric linkage to an antecedently occurring noun phrase exists over multiple sentences as in (4), then we must suppose a complete syntactic analysis of the discourse prior to the interpretation of any of its constituent sentences.

(4) A farmer owned a donkey. He beat it. It ran away.

This conclusion is not cognitively plausible and suggests that the Montague approach to anaphora is wrong. There are other difficulties too. Anaphoric dependencies across attitude contexts such as that between 'a witch' and 'she' in (5) below are also difficult to treat for Montague Grammar without assuming a *de re* reading of the attitude

reports (see further Geach 1963).

(5) Hob believes that a witch blighted his mare. Nob believes she killed his cow.

2 Dynamic semantics

Kamp (1981) and Heim (1982) independently proposed a formal and rigorous solution to the problem of anaphoric pronoun interpretation that has come to be called 'dynamic semantics'. The solution redefines the semantic contribution of a sentence to the content of a discourse. In Montague Grammar, the contribution of a sentence is a proposition - formally, a set of possible worlds in which the sentence is true (see Semantics, possible worlds). Such a proposition contributes to the content of a discourse in a simple way: the meaning of a discourse is just those possible worlds that are in the intersection of all the propositions that are the meanings of the discourse's constituent sentences. In dynamic semantics, the interpretation of a sentence S is a function from a discourse context to another discourse context. Such contexts may be understood in different ways (see Barwise 1987, Groenendijk and Stokhof 1991, Webber 1978). Sketched below is Kamp's approach to discourse semantics, 'Discourse Representation Theory' or DRT.

A discourse context in DRT contains a set of discourse entities or 'discourse referents' to which elements of subsequent discourse may refer. DRT assigns a truth-conditional meaning to a natural language discourse in two steps. First, we construct a representation of the content of the discourse known as a 'discourse representation structure' or DRS. Second, we embed a DRS in a Tarskian model of the sort familiar from first-order logic to provide truth-conditions. A DRS is naturally construed as a partial model and conceptual representation of the discourse. It consists of a pair of sets: a set of discourse referents, or universe; and a set of conditions. The universe of a DRS is analogous to the domain of a partial model; it contains the objects (discourse referents) talked about in the discourse. Conditions are property ascriptions to these discourse referents (Asher 1993). DRSs as partial models should be distinguished from the language used to describe them. The DRS language uses a box notation - the upper part of the box lists the discourse referents in the universe of the DRS, while the lower part of the box describes the conditions. For instance, the DRS for 'a farmer owns a donkey' is given below. It tells us that the discourse speaks of two entities, one a farmer (x), the other a donkey (y), and that the farmer owns the donkey.

(K2a)

$$\boxed{\begin{array}{c} x \quad y \\ \hline \\ \mathrm{farmer}(x) \\ \mathrm{donkey}(y) \\ \mathrm{owns}(x,y) \end{array}}$$

(Wada and Asher (1986), Zeevat (1989) and Asher (1993) develop a construction procedure whereby each lexical element contributes some sort of DR-theoretic structure. These then combine together compositionally following the syntactic structure of the sentence to produce a DRS for the sentence.) A noun phrase such as 'a farmer' introduces a novel discourse referent into the DRS as well as a condition on that discourse referent, while verbs introduce conditions on the discourse referents introduced by the noun phrases that constitute their syntactic arguments. Anaphoric pronouns, following the original Kamp treatment of DRT in 1981, are treated analogously to bound variables; they introduce occurrences of discourse referents that have been introduced previously by the processing of noun phrases in prior discourse.

The construction of a DRS for a discourse proceeds incrementally, exploiting the syntactic parse of each sentence. If $\mathbf{K}^{\tilde{}}$ is the DRS derived from the first n sentences and K_{n+1} is the DRS derived from the sentence $n+1$, then the DRS for the $n+1$ sentence discourse is just the DRS that combines the universes of $\mathbf{K}^{\tilde{}}$ and K_{n+1} and their condition sets:

$$\langle (\mathrm{U}_{\mathbf{K}^{\tilde{}}} \cup \mathrm{U}_{K_{n+1}}), (\mathrm{Con}_{\mathbf{K}^{\tilde{}}} \cup \mathrm{Con}_{K_{n+1}}) \rangle.$$

So in constructing a DRS for (4), for example, we would add to the DRS K2a above the conditions 'beats(x, y)' and 'runs away(y)' to get a DRS for (4):

(K4)

$$\begin{array}{|c|}\hline x \quad y \\\hline \\ \mathrm{farmer}(x) \\ \mathrm{donkey}(y) \\ \mathrm{owns}(x,y) \\ \mathrm{beats}(x,y) \\ \mathrm{runs\ away}(x,y) \\\hline\end{array}$$
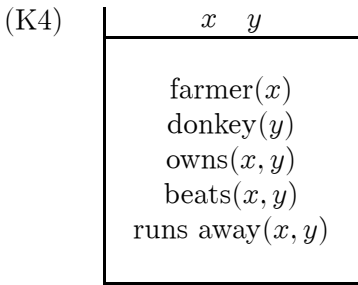
DRT solves another problem for Montague Grammar, having to do with indefinites, pronouns and conditionals. In Montague Grammar (as for Frege), the noun phrase 'a man' contributes a property of properties - that is, the property of some property P that there is some man that has P or, in symbols, $\lambda P \exists x(\mathrm{man}(x)\ \&\ P(x))$. But then how do we account for the contribution of 'a farmer' in (6)?

(6) If a farmer owns a donkey, he beats it.

If we translate 'a farmer' with the existential quantifier, we must assign it wide scope over the conditional in order to bind the variable introduced by 'he' which is intuitively linked to 'a farmer'. But this gives (6) incorrect truth-conditions. (6) requires the following translation for 'a farmer':

$\lambda P \forall x(\mathrm{farmer}(x) \rightarrow P(x))$.

Montague Grammar has no uniform translation of the indefinite determiner that yields a correct treatment of (2) and (6), and also fails to explain why its translation of indefinites is context sensitive in this way. DRT, however, has a uniform translation. DRT uses its analysis of indefinite noun phrases and pronouns, together with construction procedure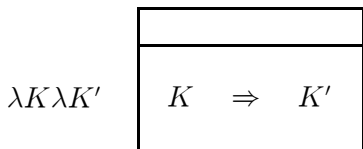s for conditionals and universally quantified NPs. Both constructions introduce what is known as a 'complex condition' in a DRS; complex conditions consist of DRSs as arguments to some operator. For instance, the grammatical conjunction 'If…, then… ' introduces the following relation on DRSs:

$$\lambda K \lambda K' \quad \begin{array}{|c|}\hline \\\hline K \quad \Rightarrow \quad K' \\\hline\end{array}$$

When we have two clauses linked by 'If…, then… ', the first clause gives us a DRS replacing the variable K, while the second gives us a DRS replacing K'. For example, if we choose the right discourse referents for the pronouns 'he' and 'it', we translate (6) as in the DRS below, using a uniform treatment of indefinites, conditionals and pronouns.

(K6)

$$\begin{array}{|c|}\hline \\\hline \begin{array}{|c|} \hline x \quad y \\\hline \mathrm{farmer}(x) \\ \mathrm{donkey}(y) \\ \mathrm{owns}(x,y) \\\hline \end{array} \Rightarrow \begin{array}{|c|} \hline \\\hline \mathrm{beats}(x,y) \\\hline \end{array} \\\hline\end{array}$$
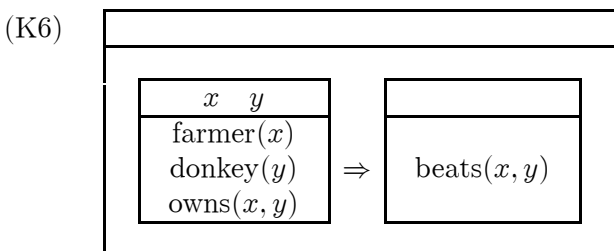
As K6 shows, DRSs can occur within DRSs. DRT postulates a constraint on anaphora called 'accessibility' which exploits these complex structures. A discourse referent ' α' is accessible from a condition ' β' if and only if ' α' is declared in the universe of a DRS that either contains ' β' or a DRS containing ' β', or is the antecedent of a conditional ' $\Rightarrow$' of which the DRS containing ' β' is the consequent. We may translate a pronoun with the discourse referent ' α' only if ' α' is accessible to the conditions of which the pronoun is the argument. This means that DRT predicts some anaphoric connections to be semantically incoherent - for example, as in 'If a farmer $_i$ owns a donkey, he $_i$ beats it. *He $_i$ is fined'.

3 Dynamic semantics (cont.)

Let us now briefly discuss the correctness definition for DRSs. The correctness definition tells us what conditions must obtain in order for a DRS to be properly embedded in a 'DRS model'. These are the truth-conditions for the discourse the DRS represents. A DRS model is a pair $\langle D, I \rangle$, where D is a non-empty set (a domain of individuals) and I is a function that assigns to atomic n-ary conditions of DRSs n-tuples from D^n. Atomic conditions are those conditions that are derived from natural language nouns, verbs and some adjectives - the sort that are contained in the DRSs for (2) and (6). A DRS K is properly embedded in a DRS model M if and only if there is a function from U_K into the universe of M such that all the conditions of K are satisfied in M. What we have to do now is to define satisfaction of a condition in K. For atomic DRS conditions, the definition of satisfaction is completely analogous to the satisfaction of an atomic formula of first-order logic relative to a model and an assignment to free variables. The recursive definition for, for example, complex conditions of the form $K \Rightarrow K'$, however, is novel, reflecting the quantificational force of '$\Rightarrow$'. $K \Rightarrow K'$ is satisfied in a DRS model M relative to an embedding function f if and only if for every function g that extends f to a proper embedding of K in M there is an extension h of g that is a proper embedding of K' in M.

In applying the correctness definition to K4, we get the intuitively right truth-conditions for the sentence. The function f is a proper embedding of K4 just in case there is an object a and an object b such that a is a farmer, b is a donkey, a owns b, a beats b, and b runs away. The correctness definition also captures the truth-conditions of a conditional sentence such as (6) - for more details see Kamp and Reyle (1993).

Since its introduction, DRT and other theories of dynamic semantics have been used to analyse various anaphoric phenomena - for example, plural anaphora (Kamp and Reyle 1993), anaphora across attitude contexts (Asher 1987), temporal anaphora (Kamp and Rohrer 1983, Kamp and Reyle 1993) and verb phrase (VP) ellipsis (Klein 1986).

DRT makes an important contribution to our understanding of anaphora. But there are problems. DRT's approach to anaphors and quantification sometimes fails to get the right readings in examples such as the following:

(12) If I have a dime in my pocket, I'll put it in the meter.

DRT predicts that every dime I have in my pocket will be put in the meter, a prediction that must somehow be blocked. Sentences involving pronouns and other quantifiers yield similar difficulties. (13) casts doubt on a basic principle of both Montague Grammar and DRT; the 'it' in (13) does not seem to function as a bound variable but rather as a definite description (Evans 1980).

(13) Either there's no bathroom in this house, or it's in a funny place.

Yet another difficulty for the DRT account of anaphora is that it cannot handle many cases of propositional anaphora. DRT does not offer for (14), for instance, any entities of the appropriate type to which 'it' can refer.

(14) One plaintiff was passed over for promotion. Another didn't get a pay increase for five years. A third received a lower wage than men doing the same work. But the jury didn't believe any of it.

One approach to propositional anaphora such as that in (14) has been to analyse in greater detail the discourse structure of a text. Interpreters naturally understand a text as divided into meaningful segments related in various ways. Some parts give a narrative; others furnish a background for a narrative; still others may elaborate or explain other parts. Such discourse structure is completely missing from DRT, but Asher (1993) and Webber (1991) argue that these parts of the text can serve as referents to anaphoric pronouns. Notions of discourse structure have been incorporated into a dynamic-semantic framework (Asher 1993) to analyse propositional anaphora, VP ellipsis and temporal anaphora (Lascarides and Asher 1993).

See also: Discourse; Logical and mathematical terms, glossary of

NICHOLAS ASHER

References and further reading

Asher, N. (1987) 'A Typology for Attitude Verbs and their Anaphoric Properties', *Linguistics and Philosophy* 10:

125-97.(This research paper gives a detailed study of anaphora across attitude contexts within DRT.)

Asher, N. (1993) *Reference to Abstract Objects in Discourse*, Dordrecht: Kluwer.(This book offers an extension of DRT for the purpose of analysing anaphoric reference to propositions and other abstract objects.)

Barwise, J. (1987) 'Noun Phrases, Generalized Quantifiers, and Anaphora', in E. Engdahl and P. Gärdenfors (eds), *Generalized Quantifiers: Linguistic and Logical Approaches*, Dordrecht: Reidel.(This article offers an alternative formulation of dynamic semantics. It does not use semantic representations such as DRSs.)

Chomsky, N. (1981) *Lectures on Government and Binding: The Pisa Lectures*, Dordrecht: Foris.(This is one of the central works on syntactic constraints on anaphora.)

Evans, G. (1980) 'Pronouns', *Linguistic Inquiry* 11: 337-62.(This is one of the classic articles in the philosophy of language on anaphora. It introduces an alternative to the bound variable approach.)

Gamut, L.T.F. (1991) *Logic, Language and Meaning*, esp. vol. 2, *Intensional Logic and Logical Grammar*, Chicago, IL: University of Chicago Press.(This book is a very good introduction to semantics. Dynamic semantics is presented in volume 2.)

Geach, P.T. (1963) *Reference and Generality*, Ithaca, NY: Cornell University Press. (This classic text includes many puzzles for semantics.)

Groenendijk, J. and Stokhof, M. (1991) 'Dynamic Predicate Logic', *Linguistics and Philosophy* 14: 39-100. (This article presents an alternative to DRT's treatment of dynamic semantics. It is similar to Barwise's approach but better worked out.)

Heim, I. (1982) *The Semantics of Indefinite and Definite Noun Phrases* (Ph.D. thesis), Amherst, MA: University of Massachusetts Press.(Besides Kamp's 1981 paper, this is the principal source for dynamic semantics. It is devoted to the study of indefinite and definite noun phrases and their behaviour as anaphoric referents.)

Kamp, H. (1981) 'A Theory of Truth and Semantic Representation', in J. Groenendijk, T. Janssen and M. Stokhof (eds) *Formal Methods in the Study of Language*, Amsterdam: Mathematisch Centrum Tracts, 277-322; repr. in J. Groenendijk *et al*. (eds) *Truth, Interpretation and Information*, Dordrecht: Foris, 1984, 1-41.(This is Kamp's first paper on DRT. It is difficult to read but is the principal source for many if not most of the developments in dynamic semantics.)

Kamp, H. and Reyle, U. (1993) *From Discourse to Logic: Introduction to Model-Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht: Kluwer.(This book is a lengthy but readable introduction to DRT. It is written for the non-specialist but also includes valuable material for the specialist.)

Kamp, H. and Rohrer, C. (1983) 'Tense in Texts', in R. Bäuerle, C. Schwarze and A. von Stechow (eds) *Meaning, Use and Interpretation of Language*, Berlin: de Gruyter, 250-69.(This article describes a dynamic-semantic approach to tense and temporal anaphora. The study is carried out in DRT.)

Klein, E. (1986) 'VP Ellipsis in DR Theory', in J. Groenendijk (ed.) *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, Dordrecht: Foris.(This paper shows how dynamic semantics can be used to analyse VP ellipsis.)

Lascarides, A. and Asher, N. (1993) 'Temporal Interpretation, Discourse Relations, and Commonsense Entailment', *Linguistics and Philosophy* 16: 437-93.(This article criticizes the dynamic-semantic treatment of tense and temporal anaphora.)

Wada, H. and Asher, N. (1986) 'BLDRS: A Prolog Implementation of LFG and DR Theory', *Proceedings of the 11th International Conference on Computational Linguistics*, Bonn, 540-5.(This was one early implementation of DRT and suggested a compositional treatment of dynamic semantics. Some other such implementations are mentioned in the bibliography.)

Webber, B.L. (1978) *A Formal Approach to Discourse Anaphora*, report 3761, Boston, MA: Bolt, Beranek & Newman Research.(This is a computational approach to anaphora. It has some similarities to dynamic semantics.)

Webber, B.L. (1991) 'Structure and Ostension in the Interpretation of Discourse Deixis', *Language and Cognitive Processes* 6: 107-36. (This article investigates a variety of anaphoric phenomena in discourse.)

Zeevat, H. (1989) 'A Compositional Approach to Discourse Representations', *Linguistics and Philosophy* 12: 95-131.(This article gives the first compositional treatment of DRT that brings out the similarities with dynamic predicate logic.)

Animal language and thought

The question of animal language and thought has been debated since ancient times. Some have held that humans are exceptional in these respects, others that humans and animals are continuous with respect to language and thought. The issue is important because our self-image as a species is at stake.

Arguments for human exceptionalism can be classified as Cartesian, Wittgensteinian and behaviourist. What these arguments have in common is the view that language and thought are closely associated, and animals do not have language. The ape language experiments of the 1960s and 1970s were especially important against this background: if apes could learn language then even the advocates of human exceptionalism would have to admit that they have thoughts. It is now generally believed that whatever linguistic abilities apes have shown have been quite rudimentary. Yet many sceptics are willing to grant that in some cases apes did develop linguistic skills to some extent, and clearly evidenced thought. Studies of other animals in captivity and various animals in the wild have provided evidence of highly sophisticated communicative behaviour. Cognitive ethology and comparative psychology have emerged as the fields that study animal thought. While there are conceptual difficulties in grounding these fields, it appears plausible that many animals have thoughts and these can be scientifically investigated.

1 Human exceptionalism versus continuity across species

Richard Sorabji (1993) has argued that debates about animal language and mind go to the core of the western philosophical tradition. Aristotelians and Stoics argued that only humans have reason or belief; some Platonists and Pythagoreans argued that these are shared by many kinds of animal. Indeed, Plato himself challenged the very framework presupposed by the debate. He thought that it was just as sensible to divide the world into cranes and non-cranes as humans and non-humans (see *Statesman* 263d).

Both human exceptionalism (HE) and continuity across species (CAS) have had strong supporters. HE was defended by Aquinas, Descartes and many twentieth-century linguistic philosophers, CAS by Voltaire, Hume and Darwin.

Although it is easy to characterize HE and CAS generally, it is difficult to do so precisely. Roughly, those who espouse HE believe that humans are unique in having language and sophisticated thought, and that there is a deep chasm between these human capacities and whatever thoughts and communication systems other animals may have. Those who embrace CAS hold that humans share the capacity for language or thought with at least some other animals and that the differences with respect to these capacities between humans and other animals are gradual and incremental.

However advocates of HE need not hold that humans are exceptional with respect to every capacity and every animal. They may contemplate honorary humanhood for dolphins (for example), or grant that a few other animals have thoughts but insist that these thoughts are always first-order or nonconscious, and thus very different from the thoughts that humans are capable of having. Defenders of HE may allow that many animals have communication systems, but then go on to claim that these are vastly weaker and less sophisticated than human language.

Supporters of CAS may grant that humans are capable of having some thoughts that no other animal can have. But they will generally see this as an evolutionary fact about humans that is not importantly different from other evolutionary facts about other animals. Perhaps only humans can ponder the neurophysiology of the wildebeest, but lions can think thoughts about wildebeests that humans cannot conceive. Defenders of CAS sometimes grant that only humans use language, but they often see this as a matter of definition or otherwise trivial. They are inclined to see complex communication systems as similar to languages or as sophisticated and important. Although the difference between HE and CAS may be vague, proponents of HE assert that there are enormous differences between humans and animals that centre on language and thought while advocates of CAS deny this.

The dispute between HE and CAS is important for several reasons. If animals lack central features of language and thought that humans have, then a profound gulf separates us from them. The existence of this gulf may have implications about the relations between the natural sciences and the human sciences. It may also justify discontinuous moralities with respect to humans and animals (see Animals and ethics; Moral standing). If HE is

correct, then we may be justified in seeing ourselves as special - perhaps even as 'the crown of creation'. On the other hand, if CAS is correct, this may mean that action theory and philosophy of language should be seen as branches of ethology, and that our treatment of animals is a moral scandal. We may have to give up the view of ourselves as morally and metaphysically privileged, and instead see ourselves as one animal species among many. What is potentially at stake in arguments about animal language and thought is our human self-image - who we are, what we are like and what constitutes our proper relations with the rest of nature.

2 Arguments for human exceptionalism

Many philosophers have defended HE. Any attempt to collect these views into categories and to develop generic arguments involves regimentation. With these caveats in mind, it is useful to divide the arguments for HE into three categories: Cartesian, Wittgensteinian and behaviourist.

Cartesian views about animal language and thought have been influential on philosophers and linguists such as Vendler (1972) and Chomsky (1966). Although there is controversy about the exact nature of Descartes' views about animals, the broad outlines are clear.

Chomsky credits Descartes with recognizing that language use is 'creative': it is both unbounded in scope and stimulus-free. Descartes wrote that while 'magpies and parrots are able to utter words just like ourselves' this is mechanical, 'a movement of mere nature' rather than a sign of thought (*Discourse on Method*, 1637: Part V). Having established to his satisfaction that animals do not have language, Descartes infers that they do not have thought, 'for the word is the sole sign and the only certain mark of the presence of thought hidden and wrapped up in the body' (letter to Henry More, 1649). Although Descartes is ambivalent about whether it can be proved that animals do not have thoughts, clearly he believes that they do not. Since animals do not have thoughts they do not have 'real feeling', for 'real feeling' involves propositional content and animals are incapable of propositional content because they do not have language.

Wittgensteinian accounts of animal language and thought have been given by Malcolm (1972-3) and Leahy (1991). Wittgenstein's own views are characteristically difficult to unravel. In *Philosophical Investigations* (1953) he claims that animals 'do not use language - if we except the most primitive form of language', but appears to think that animals have sensations, emotions, intentions and first-order beliefs. However, Wittgenstein denies that animals have the power to simulate pain, to talk to themselves, or have attitudes about future events.

According to Malcolm, animals think but do not have thoughts. Having thoughts involves formulating and entertaining propositions, and he believes that animals are incapable of this. Although Malcolm (1972-3) does not identify thought with language, he claims that the relationship is 'so close that it is really senseless to conjecture that people may *not* have thoughts, and also really senseless to conjecture that animals *may* have thoughts'. However, we know that animals think because 'in real life we commonly employ the verb "think" in respect to animals'. Clearly Malcolm believes that animal thinking does not involve having thoughts, but says very little about how we are to understand it.

Behaviourists in both science and philosophy have denied animal language and thought (see Behaviourism, analytic; Behaviourism, methodological and scientific). Although some behaviourists deny thought to humans as well, most reserve their deepest scepticism for animals. Quine (1960) takes it as obvious that animals do not have language, and that ascribing thoughts to animals is an 'essentially dramatic idiom' - we imagine ourselves in the animal's place and say what thoughts we would have were we the animal.

Davidson (1975) has produced the following argument for supposing that animals do not have thoughts: if an animal has a thought, then this thought must occur in a network of beliefs. This follows from Davidson's holism which he takes from Quine: thoughts or beliefs come in 'webs', they cannot occur singly. In order to have a network of beliefs, an animal must have the concept of belief. This is because having a network of beliefs requires the ability to distinguish between someone holding a sentence to be true and the sentence actually being true. But having the concept of belief requires having language, for Davidson believes that this concept only arises in the context of linguistic interpretation. Since no animals have language they do not have the concept of belief. Since they do not have the concept of belief they do not have networks of beliefs. Hence, they do not have beliefs at all (see Davidson, D. §8).

What is striking about the arguments for HE, taken together, is that they turn on supposing a very close connection between language and thought. All of the arguments that we have reviewed suppose that having language is a necessary condition for having thoughts (although Malcolm grants that some kind of thinking may occur in the absence of language and thought). Moreover, some of these philosophers (for example, Malcolm) think that having language is sufficient for having thoughts.

It is against this background that the ape language experiments of the 1960s and 1970s caught the attention of philosophers. If it could be shown that apes could learn language then many philosophers would be convinced that apes are capable of having thoughts. Indeed, some philosophers would be convinced by nothing short of this.

3 The ape language controversy

Since the beginning of the twentieth century there have been at least half a dozen attempts to teach spoken language to an ape. In 1966 Beatrice and R. Allen Gardner began teaching American Sign Language to a chimpanzee named Washoe, and throughout the late 1960s and 1970s studies employing animals of different species using different communication systems were undertaken by David Premack, Duane Rumbaugh, Roger Fouts, Francine Patterson and others. In some cases extravagant claims were made about the linguistic abilities of these apes. These claims were deflated by Herbert Terrace in the late 1970s.

From 1973 to 1977 Terrace studied a chimpanzee named Nim Chimpsky. He concluded (1980) that there is no unequivocal evidence for supposing that apes can master syntactic, semantic or pragmatic aspects of language. Terrace's results were devastating to the credibility and funding base of the ape language projects. The projects of Patterson and Fouts continue, funded primarily by private donations. The Rumbaugh project, now headed by Sue Savage-Rumbaugh, is one of the few that continues to produce significant scientific publications.

Beginning in 1981 Savage-Rumbaugh turned her attention to bonobos (so-called 'pygmy chimpanzees'). Using a specially designed keyboard connected to a speech synthesizer one bonobo, Kanzi, has shown a surprising understanding of spoken English. He has a large vocabulary, is capable of communicating novel information and clearly follows simple syntactic rules. In a comparative study (1994) Savage-Rumbaugh concluded that the eight-year-old Kanzi's linguistic skills were superior to those of a normal human two-year-old.

The ape language studies have been controversial from the beginning and continue to be so today. Terrace and other critics accuse most researchers of inaccurate observation and analysis, overinterpretation and cuing desired behaviour. Terrace has been faulted for failing to obtain results because of a lack of rapport with his subject and an impersonal training regime.

Animal behaviour, like human behaviour, is indeterminate, description-relative and open to interpretation. John Dupré (1991) has argued that the dispute over the ape language experiments primarily involves conflicts about the goals and methods of scientific research. The critics demand that claims about the linguistic abilities of apes be backed by compelling evidence that conforms to the most rigorous canons of scientific methodology. However, in the case of humans, language develops in a highly complex and emotive context and the application of principles of charity is an important part of the language-learning process (see Charity, principle of §4). It may be that there is an intrinsic conflict between teaching apes language in the most effective way possible and doing so in a way that will satisfy the scientific scruples of the sceptics.

It is clear that the ape language experiments have not convinced philosophers that at least some animals have language and therefore thoughts. However, one point is worth considering. Even sceptics may admit that some apes have demonstrated linguistic capacities to some degree or in some respect. But a similar view with respect to having thoughts is difficult to even understand. Having a thought appears to be an all or nothing matter. This may provide some evidence against the view that language and thought are as closely tied as some philosophers have claimed.

4 Animal communication

Even some scientists and philosophers friendly to the idea that animals have thoughts have been sceptical of the ape language experiments. For these experiments have focused on very few individuals from an even smaller number of species. They have been directed towards eliciting a variety of behaviours that these animals do not

manifest under natural conditions.

In addition to the ape language experiments there have been other interesting studies of the communication abilities of other captive animals (see Bekoff and Jamieson 1995; Ristau 1991). Louis Herman has claimed that bottle-nosed dolphins are capable of semantic and syntactic processing that 'utilize a rich network of mental representations when responding to language-mediated tasks', although they are less able to produce semantically and syntactically dense utterances. Irene Pepperberg has trained an African grey parrot to respond accurately to questions about the colour, shape, name and category of a variety of objects.

Under natural conditions many animals engage in highly sophisticated communicative activities. Cheney and Seyfarth have shown that vervet monkeys distinguish several different kinds of alarm call and behave appropriately with respect to each. Carolyn Ristau suggests that the broken wing displays in various species of plovers may be intended to lead intruders away from their nests. Deception has been claimed for a variety of animals including chimpanzees and elephants.

5 Thought without language

As we have come to know more about animals, the possibility that they may have thoughts without language has become increasingly plausible to many people. Even Joel Wallman (1992), one of the most thorough-going critics of the ape language experiments, holds that apes are 'highly intelligent', 'reflexive' and capable of 'impressively abstract mentation'. Cognitive ethology and comparative psychology have emerged as the scientific fields that study the cognitive capacities of animals that may underlie some of their behaviour.

The dominant view in cognitive ethology, as well as among some philosophers who have written about cognitive ethology, is that an animal's cognitive states must be inferred from its behaviour. Since an animal's mental states are never observed but only inferred, the claim that an animal has a particular mental state or cognitive capacity can only be more or less probable. Often it is said even by cognitive ethologists that we can never really know that animals have thoughts.

In my view, this way of approaching the question of animal thought is fundamentally flawed. If by hypothesis the existence of animal thoughts can only be known by inference from behaviour, then it will always be an open question whether or not animals have thoughts. But if there are no convincing conceptual arguments for supposing that they do not, then we are justified in approaching the question of animal thoughts in the way that we approach the question of human thoughts. For there is no behaviour that an animal can engage in that will ineluctably drive us to the conclusion that it has thoughts. But if we were to view the claim that humans have thoughts as an inference from human behaviour, a sceptic could consistently deny the conclusion (see Other minds). As a matter of fact, however, we learn to view other humans as having thoughts in the process of psychological and social development. In many cases it is true to say that we see their thought in action. We can be wrong about the content of a particular thought, and indeed some versions of solipsism and scepticism may be logically possible or even defensible in a philosophy journal or seminar. But we do not confuse this with the practices of everyday life. Similarly, it seems that viewing some animals as sometimes having thoughts is part of our cultural practice. While we can take a sceptical stance and it may be useful to do so for certain methodological purposes, there is little reason to do so when it comes to fixing our beliefs. The idea that we are continuous with animals, unless it can be shown otherwise, appears to be a plausible naturalistic hypothesis.

Some philosophers may still be bothered by the fact that animals do not have full-blown human languages. But many animals may have powerful enough, nonlinguistic, representational systems to make their thoughts a significant possibility. For that matter, philosophers such as Ryle who deny that thinking necessarily involves manipulating representations may well be correct. At this stage in the investigation, the possibility that animals may have thoughts without language does not appear to be blocked by any convincing philosophical argument.

DALE JAMIESON

References and further reading

Bekoff, M. and Jamieson, D. (eds) (1995) *Readings in Animal Cognition*, Cambridge, MA: MIT Press.(This is a collection of articles on animal cognition by leading philosophers, psychologists and biologists, including Cheney and Seyfarth, Herman, Ristau and Savage-Rumbaugh.)

Chomsky, N. (1966) *Cartesian Linguistics*, New York: Harper & Row.(A defence of the Cartesian foundations of Chomsky's linguistic theory.)

Davidson, D. (1975) 'Thought and Talk', repr. in *Truth and Interpretation*, Oxford: Oxford University Press, 1984, 155-70.(This is the most influential recent article that denies that animals have thoughts.)

Dupré, J. (1991) 'Conversations with Apes: Reflections on the Scientific Study of Language', in J. Hyman (ed.) *Investigating Psychology: Sciences of the Mind After Wittgenstein*, London: Routledge, 95-116.(This is an especially interesting philosophical assessment of the ape language studies.)

Jamieson, D. (1998, forthcoming) 'Science, Knowledge, and Animal Minds', *Proceedings of the Aristotelian Society*.(Discussion of the 'other minds' problem as applied to animals.)

Leahy, M.P.T. (1991) *Against Liberation*, London: Routledge.(This attack on animal liberation argues that talk about animal thought and language is largely anthropomorphic.)

Malcolm, N. (1972-3) 'Thoughtless Brutes', repr. in D.M. Rosenthal (ed.) *The Nature of Mind*, New York: Oxford University Press, 1991, 454-61.(A Wittgensteinian defence of the view that animals think but do not have thoughts.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Foundational work in behaviourist philosophy of language.)

Ristau, C. (ed.) (1991) *Cognitive Ethology*, Hillsdale, MI: Lawrence Erlbaum Associates.(A collection of essays including the work of Ristau, Cheney and Seyfarth, and Pepperberg.)

Savage-Rumbaugh, E.S. and Lewin, R. (1994) *Kanzi: The Ape at the Brink of the Human Mind*, New York: Wiley.(The popular exposition of Savage-Rumbaugh's work.)

Savage-Rumbaugh, E.S. *et al.* (1993) 'Language Comprehension in Ape and Child', *Monographs of the Society for Research in Child Development* 233 (58).(The technical exposition of Savage-Rumbaugh's work. Co-authors are Jeanine Murphy, Rose Sevcik, Karen Brakke, Shelly Williams and Duane Rumbaugh, with commentary by Elizabeth Bates.)

Sorabji, R. (1993) *Animal Minds and Human Morals*, Ithaca, NY: Cornell University Press.(The best discussion of the ancient debate over animal minds.)

Terrace, H.S. (1980) *Nim*, London: Methuen.(A critique of the ape language experiments and the report of a failed attempt to teach language to a chimpanzee.)

Vendler, Z. (1972) *Res Cogitans*, Ithaca, NY: Cornell University Press.(A contemporary defence of a Cartesian theory of mind.)

Wallman, J. (1992) *Aping Language*, New York: Cambridge University Press.(A systematic, sceptical discussion of the ape language controversy.)

Wittgenstein, L. (1953) *Philosophical Investigations*, ed. G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, Oxford: Blackwell.(Profoundly influential discussion of language and mind.)

Communication and intention

The classic attempt to understand communication in terms of the intentions of a person making an utterance was put forward by Paul Grice in 1957. Grice was concerned with actions in which a speaker means something by what they do and what is meant might just as much be false as true. He looked for the essence of such cases in actions intended to effect a change in the recipient. Grice saw successful communication as depending on the recognition by the audience of the speaker's intention. Since then there have been many attempts to refine Grice's work, and to protect it against various problems. There has also been worry that Grice's approach depends on a false priority of psychology over semantics, seeing complex psychological states as existing independently of whether the agent has linguistic means of expressing them.

1 Grice's theory

The classic attempt to understand communication in terms of the intentions of a person making an utterance was put forward by Paul Grice in his 1957 article. Grice takes as his subject what he calls 'non-natural meaning' (meaning$_{NN}$), which he contrasts with the wider class of cases of natural meaning in which we say, for instance, that 'those spots mean measles' or 'those clouds mean rain'. It cannot be true that those spots mean measles but the patient has not got measles; meaning is simply a correlation or signalling. Whereas Grice is concerned with actions in which a speaker (or, in general, an agent) means something by what they do and what is meant might just as well be false as true. Here too he makes a distinction: between a wider class of actions that predictably cause a recipient to believe something, and a narrower class of actions that not only may cause the belief that p, but also themselves mean that p. The wider class would include such actions as leaving photographs around, expecting the recipient to believe that the events shown had happened. The narrower class should take us somewhere near the domain of linguistic communication.

Grice looks for the essence of such cases in actions intended to effect a change in the recipient. He eventually located the difference in the manner in which the change in belief is expected to come about. In the wider class, the action may simply be a natural sign of the state of affairs, but in the narrower class of non-natural meaning there is an element of reflexivity involved, since the intended mechanism, through which the recipient is expected to come to the belief, is that the agent is recognized as acting with the intention that the recipient come to believe it: it is only if the agent's intention is recognized that the change in belief will ensue. So, for example, deliberately ignoring someone in the street might be said to mean$_{NN}$ that the agent is angry with the recipient, because getting the recipient to believe this depends upon their recognizing the intention with which the action was performed. Grice proposes that 'A meant$_{NN}$ something by x' is roughly equivalent to 'A uttered x with the intention of inducing a belief by means of the recognition of this intention'. He sees successful communication as requiring recognition of the speaker's intention in communicating. Of course, the account could be extended to speech acts other than assertion: for example, intentions to get people to perform actions or supply information can be used to fix what is meant by imperatives and questions.

Armed with an account of the speaker's meaning on an occasion, Grice goes on to propose that 'the utterance x meant something' is roughly equivalent to 'someone meant something by x'. For a sentence to mean something 'timelessly' is for people in general to use it to mean that thing. We can thus progress from an account of communication in a 'one-off' case to a more general account of the kind of meaning possessed by the sentences of a public language.

2 Reactions

Grice's short article spawned a huge literature, and various modifications were suggested. One is that it may not be quite right to identify the speaker's intention as one of changing the hearer's belief. I may say that p without caring whether my hearer accepts or indeed already believes that p. A change in belief might be one effect of my utterance, but it is not always one I intend (in the terms introduced by Austin, it is a 'perlocutionary' effect), whereas we are looking for a characterization of what a speaker is doing in uttering anything (the 'illocutionary' act; see Speech acts §1). It seems rather that the only reliable intention behind an utterance is that the speaker intends the hearer to understand something by it, not necessarily to believe it or react to it in any other

way. But already it is a matter of fine judgment whether this is important, and, after all, one can say things with the intention that the hearer not even understand them, for example when showing off one's grasp of some technical term. Certainly, a defence of Grice would go, once the whole social practice of communication with language is in place, we can say things with the most devious of intentions. Nevertheless, it may be true that the central or paradigm case is that of imparting information, and indeed any account that fails to put this function of language at the centre risks distortions of its own. So Grice may simply be following the proper path by concentrating upon the central or basic case and hoping that deviations will become explicable in the light of what is said about it.

A more elaborate debate queried whether Grice's conditions are sufficient for meaning. It concentrates on cases where there is an element of deceit. As Bennett puts it, 'in real communication everything is open and above board' (1976: 126), but a range of cases suggest that Grice's conditions could be met while there are hidden and devious intentions around. Grice's conditions could be satisfied while, further up the hierarchy of intentions, I have the intention that you misunderstand something about the situation. Complex counterexamples of this structure were rapidly concocted. However, what Grice called 'sneaky intentions' can be blocked by a number of strategies. One such is to require that all the speaker's intentions be known to the hearer, and another is that the intentions in question should be 'mutually known', where this means that the hearer knows the speaker's intentions, the speaker knows that the hearer knows, the hearer knows that the speaker knows…and so on.

There is a cost to piling up the reflexive intentions with which speakers are credited, for it is easy to wonder whether there is a psychological reality to these complex layers of intention, or whether they are simply an artefact of the theory. In fact, some unease arises when we realise that Grice's approach requires that the reflexive intention - the intention that the hearer's belief be modified by means of their recognition of my intention in speaking - be typically present throughout the field of linguistic communication. It would seem to be to overload the psychology of the young child or not-very-aware adult to credit them with intentions of such complexity: typical speakers, we might say, just do not know or care *how* their messages get across, so long as they do. Certainly a strategist faced with a hitherto unknown type of problem of communication may do something with the hope that the intention with which they do it is recognized, but this seems to be a sophisticated plan for coping with precisely the kind of situation where normal communication has broken down.

This kind of thought suggests that Grice is too much concerned with one-off cases. At least in typical cases of communication we are not involved in one-off strategies, but can rely upon the conventional meanings of our terms. (For Grice himself, as we have seen, the existence of these conventions would simply be the existence of enough people inclined to use a term to effect communication in enough cases.) While reflexive intentions may be necessary to solve new problems of communication, as when we confront someone with whom we share no language, it may be that they drop out of the picture when a rule-governed or conventional medium of communication is available: it may be that reliance on conventions and rules supplants the mechanism of recognizing the speaker's intention, rather than supplementing it. Once we have language I do not have to know even whether you intended me to understand you in order to know what you said. Similarly, there may be sentences whose typical use is not to convey the messages that their strict and literal meanings suggest, but nevertheless, when uttering one of them, one can be held to have said whatever it is that its strict and literal meaning identifies. In short, a conventional and deeply normative system (that is, one in which the speaker is *liable* to be held to have said things, regardless of what was intended) takes on a life of its own, independent of the detailed case-by-case intentions of participants.

Behind this objection there lies the deeper unease that Grice's approach sets things up so that the speaker has intentions of considerable complexity, and we then try to understand meaning and linguistic communication given so much rich psychology. Whereas a permanent strand in modern philosophy of language has been to try to identify the mastery of language with the ability to have thoughts (or, at least, thoughts of this kind of complexity) at all. From this standpoint it may seem perverse arbitrarily to enrich the psychology of the actors, and on that basis to explain their use of sentences as vehicles of their given, antecedent intentions. You cannot intend, the thought goes, without representing to yourself what you intend, and typically this will be by using the best representative medium known to us, which is our natural language. In forming an intention we typically speak to ourselves, saying that we will do this or that, and if this process is itself seen as a kind of internal communication then intentions are not suitable for giving an analysis of communication in general. So we should be thinking of social communication as one of the activities that makes possible the enriched psychology that we find in

participants in conversational exchanges. The problems in this area are, however, deeply intractable. Some philosophers, notably John Searle (1983), insist that the 'intentionality' of language (that is, its directedness or power to represent absent or merely possible states of affairs) must be understood as being derived from a more fundamental power of the 'mind/brain'; in such a scheme there is no problem with Grice's direction of attack. To others it merely mystifies things to postulate intentional powers in the mind/brain, and our capacity to represent absent states of affairs to ourselves must be seen as essentially linguistic, since language is the only representational system that we actually know anything about.

A final radical criticism is that Gricean proposals try to secure openness in communication by adding to the speaker's intentions. But, it has been argued (Meijers 1994), openness cannot be secured that way. It is a matter of a relationship between the speaker and hearer, not a one-way matter of the speaker having sufficiently open intentions. A speaker must not only have the intention to be open, but there must also be a commitment to being open (and liability to penalty if they are not). Arguably this takes us into the domain of collective intentions; ones based on an understanding between the different parties about what they, collectively, are trying to do. The relation between speaker and hearer, by such an account, is more like that between us when we together plan that *we* sing a duet, than when I unilaterally intend that *I* do something with some effect on you. Again, there are various proposals about the nature of such common intentions, and whether they reduce to a kind of aggregate of individual intentions.

3 Prospects and alliances

Sir Peter Strawson (1971) detected a 'Homeric opposition' between followers of Grice and philosophers of language, such as Davidson, who put the notion of a truth-condition at the centre of their picture. It is fair to say that this opposition has softened in the intervening years. Clearly there is no formal contradiction between supposing that a remark in the mouths of members of some community means that p if and only if there exists a practice of using it with the intention of communicating that p, and saying that the same remark means that p if and only if it represents the state of affairs that p, or has the truth-condition that p, or would be interpreted by an ideal interpreter as saying that p. The opposition is a question of whether we give priority to representation over communication, or vice versa. If we think of sentences possessing meaning by way of some kind of accord with things and facts, then their role in communication will seem a kind of bonus; if we think of the seamless way in which linguistic behaviour is woven into communication (and other activities involving the world) then we will reject any appeal to basic representative powers of elements of the 'mind/brain', in favour of a clearer understanding of the representative powers of the whole person, here meaning the linguistically active, socially embodied person communicating with others of the same nature.

The perennial importance of Grice's work is that intention and communication cannot be separated for long. If I intend to tell you that the cat is sick, we have not communicated unless this is how you take my utterance. If I intend a remark ironically or condescendingly, then we have not communicated unless you understand the irony or the condescension. And there is a limit to the possibility of unintended meaning: although on an occasion someone may mean something by an utterance that they did not intend, or fail to mean something that they did intend, such occasions are essentially parasitic. If they occur too often, then the conventions shift, and the meaning of the remark in the language realigns itself with the way the speakers intend it. There is no separating semantics and psychology, but there is the perennial question of priority. The strange and intriguing interdependency of intention and language seems to be two-way: our words can convey no more than we intend by them, but we ourselves can intend only what our words will carry.

Further work on these topics clearly requires a better understanding of whether meaning is essentially social. Work on rule-following has sometimes issued in the view that it must be so, and that determinate meaning only emerges in a fully social, normative practice in which the applications of words are routinely subject to criticism and correction. If this kind of thought can finally be defended, then Grice's programme will have been thoroughly vindicated, in its direction if not in every detail.

See also: Meaning and communication; Meaning and rule-following

<div align="right">SIMON BLACKBURN</div>

References and further reading

Avramides, A. (1989) *Meaning and Mind*, Cambridge, MA: MIT Press.(A thorough investigation particularly of the issue of reduction of semantics to psychology, and whether Grice's approach requires belief in it.)

Bennett, J. (1976) *Linguistic Behaviour*, Cambridge: Cambridge University Press.(A lively development of Grice's work, especially valuable for its discussion of the relation between full intentionality and more primitive signalling systems.)

Grice, P. (1957) 'Meaning', *Philosophical Review* 66: 377-88; repr. in *Studies in the Way of Words*, Cambridge, MA, and London: Harvard University Press, 1989.(The seminal article in the field.)

Lewis, D.K. (1969) *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.(The classic modern analysis of the concept.)

Meijers, A. (1994) *Speech Acts, Communication and Collective Intentionality*, Utrecht: Leiden University.(An impressive defence of the ineliminable presence of collective intentions in communication.)

Searle, J.R. (1969) *Speech Acts*, Cambridge: Cambridge University Press.(A lucid development of J.L. Austin's work on the actions we perform in conversation.)

Searle, J.R. (1983) *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press. (Searle's work is perhaps the most impressive development of the notion of acts performed in speech, but also controversial through his firm belief that the intentional powers of language are derived from a prior biologically engendered capacity.)

Strawson, P.F. (1971) 'Meaning and Truth', in *Logico-Linguistic Papers*, London: Methuen.(An early reflection upon the relations between communication and representation.)

Communicative rationality

The concept of 'communicative rationality' is primarily associated with the work of the philosopher and social theorist Jürgen Habermas. According to Habermas, communication through language necessarily involves the raising of 'validity-claims' (distinguished as 'truth', 'rightness' and 'sincerity'), the status of which, when contested, can ultimately only be resolved through discussion. Habermas further contends that speakers of a language possess an implicit knowledge of the conditions under which such discussion would produce an objectively correct result, and these he has spelled out in terms of the features of an egalitarian 'ideal speech situation'. Communicative rationality refers to the capacity to engage in argumentation under conditions approximating to this ideal situation ('discourse', in Habermas' terminology), with the aim of achieving consensus.

*Habermas relies on the concept of communicative rationality to argue that democratic forms of social organization express more than simply the preferences of a particular cultural and political tradition. In his view, we cannot even understand a speech-act without taking a stance towards the validity-claim it raises, and this stance in turn anticipates the unconstrained discussion which would resolve the status of the claim. Social and political arrangements which inhibit such discussion can therefore be criticized from a standpoint which does not depend on any specific value-commitments, since for Habermas achieving agreement (*Verständigung*) is a 'telos' or goal which is internal to human language as such. A similar philosophical programme has also been developed by Karl-Otto Apel, who lays more stress on the 'transcendental' features of the argumentation involved.*

1 Instrumental and communicative rationality: Habermas and the Frankfurt School

Jürgen Habermas, the leading thinker of the second generation of the Frankfurt School, introduced the concept of communicative rationality in order to correct what he perceives to be the 'normative deficit' in the work of earlier thinkers in this tradition. This deficit consists in the lack of any philosophically perspicuous grounding for their critique of modern society. The thought of the Frankfurt School before Habermas, classically summarized in Theodor Adorno and Max Horkheimer's *Dialectic of Enlightenment* (1944), viewed social relations within the most advanced capitalist societies as being almost entirely shaped by the demands of 'instrumental reason'. Strongly reminiscent of Max Weber's 'purposive rationality' (*Zweckrationalität*), the concept of instrumental reason refers to the capacity to maximize efficiency in the control of objective processes through a knowledge of the determinants of such processes. For Adorno and Horkheimer the predominant institutionalization of this *aspect* of reason, reflected in its pervasive philosophical equation with reason as such, has socially disastrous consequences. It leaves the ends to be collectively pursued beyond the scope of rational determination, and gives rise to a 'totally administered society'. At the same time however, as Habermas repeatedly complains, the earlier Critical Theorists propose no alternative broader conception of reason, from the standpoint of which the restriction of reason to its instrumental aspect could be coherently criticized (see Frankfurt School; Critical theory).

The contrast between instrumental and communicative rationality, which represents Habermas' attempt to remedy this situation, is anchored in the philosophical anthropology systematically presented in *Knowledge and Human Interests* (1971). Here Habermas argues that any society, in order to reproduce itself, must be capable both of productive exchange with nature (in the form of labour), and of the communicative coordination of collective activities. Labour gives rise to concepts which articulate a 'technical interest' in instrumental control, whereas the need for agreement generates the distinct categorial framework of a 'practical interest' in hermeneutic understanding. In later writings Habermas specifies that participants in 'communicative action' must be capable of regarding the statements raised in the course of discussion from the standpoint of their *validity*, and must possess an implicit knowledge both of the appropriate procedures for settling disputed validity-claims and of the (invariably counterfactual) conditions under which following such procedures would produce an objectively correct result. Such knowledge is mobilized, for example, when discovery of covert coercion renders a previously achieved consensus invalid.

In the Western philosophical tradition, the concept of rationality has long been connected with the ability to reflect on and give grounds for one's beliefs and the actions they inform. However, in modern philosophy, it is often only beliefs capable of guiding goal-oriented action which are considered as candidates for rationality. Against this,

Habermas argues that the type of action oriented towards reaching agreement in language is irreducible to that which is oriented towards successful intervention in the objective world. The illocutionary aims of speech acts can only be achieved through cooperation, based on a free acceptance by others of the validity-claims raised by the speaker, which cannot be reduced to a causally producible effect. Hence the ability to achieve consensus by offering grounds reflects a distinct form of rationality which 'inhabits everyday communicative practice', and which Habermas therefore describes as 'communicative rationality'.

With regard to the problem of grounding a critical social theory, two aspects of Habermas' proposal are significant. First, Habermas extends the scope of validity beyond 'truth' (*Wahrheit*). He argues that claims to normative rightness (*Richtigkeit*), typically moral or legal claims, are in principle susceptible of the same form of resolution through discussion as cognitive claims. There can thus be an equivalent of truth in practical matters, although Habermas admits that the status of a claim to 'sincerity' (*Wahrhaftigkeit*) cannot be resolved through discussion. (An additional claim, concerning the 'intelligibility' (*Verständlichkeit*) of one's utterance, is only discussed sporadically by Habermas, and is perhaps best considered as a *precondition* of these principal validity-claims.) Second, canons of argument are themselves not merely relative to specific cultural and institutional contexts. Whenever we engage in argumentation, Habermas proposes, we must implicitly assume (however counterfactually) that the conditions of an 'ideal speech situation', in which the 'unforced force of the better argument' would indeed ultimately triumph, have been fulfilled, otherwise discussion would lose its point. Such a speech situation would be characterized by the equal right of all participants to raise issues, ask questions, pose objections and so on, and therefore provides a normative yardstick against which current decision-making procedures can be critically assessed. In particular, the widespread dominance of the functional requirements of 'social systems' such as the market economy and modern bureaucracy can be seen as embodying a one-sided 'rationalization' which suppresses the rational potential of those democratic principles which are equally fundamental to modernity.

2 Objections to Habermas' approach

The concept of communicative rationality faces objections from a variety of quarters. If such rationality is to represent more than the ability to adhere to culturally specific canons of argumentation, then Habermas must show the universal necessity of those pragmatic presuppositions of communication which form the ideal speech situation. The first step towards this is to argue for an internal relation between meaning and validity. Habermas contends that ' we understand a speech-act when we know what makes it acceptable' ([1981] 1984: 400), where acceptability refers to intersubjective recognition of validity and not to correspondence with a reality whose structure would supposedly be given prior to language. He then draws the conclusion that ' an orientation towards the possible validity of utterances belongs not just to the pragmatic conditions of reaching understanding, but to the understanding of language as such' (1988: 76).

However, Habermas' account of the preconditions for grasping linguistic meaning may be implausibly strong. Herbert Schnädelbach (1991), for example, has objected that to understand what makes an utterance acceptable does not require one to take a stand on whether it is, in the present instance, valid. We can observe communicative action - the exchanging of reasons - from a third-person perspective without being drawn into the process of judging validity-claims, so that a normative standpoint is not internal to the very application of the concept of communicative rationality. In reply, Habermas has argued that even in apparently unfavourable cases, such as a sudden burst of laughter during a speech, we cannot understand the laughter without knowing if it is genuine or not (and thus a valid 'utterance'), and this involves *assessing* the grounds for it. The methodological suspension of our spontaneous yes/no attitude towards validity-claims merely indicates that the task of interpretation is being postponed.

A further major objection to Habermas' views has been raised by Albrecht Wellmer, among others. It is that there may indeed be certain idealizations which function as necessary pragmatic presuppositions of linguistic communication, but that the notion of an ideal speech situation represents an 'objectivistic misinterpretation' of these. This is because an ideal speech situation, exempt from the fallibility, opacity and temporality of human language, would be beyond the conditions which make communication itself necessary and possible. To regard the striving for such a situation as morally incumbent upon us would be to regard us as obliged to attain the impossible, which is absurd. According to Wellmer (1986), communication indeed relies upon 'performative

idealizations' (for example, the conviction that the reasons we have for a claim are sound and will stand the test of time), but to transform these idealizations into an ideal to be practically pursued is to fall prey to a 'dialectical semblance' (*Schein*). Habermas, however, continues to maintain that it is meaningful to idealize forms of communication, just as we idealize physical measurements; in other words, to anticipate the full realization of conditions we must already presume to hold to some degree when engaged in serious argumentation.

Other critics have focused on Habermas' claim that reaching an understanding can be seen as the primary function of language, to which we are implicitly committed as soon as we speak. Thinkers influenced by post-structuralism, in particular, have objected that the commitment to achieving consensual truth cannot, except by begging the question, be considered more fundamental than playful, ironic or fictional uses of language. More generally, it is contended that an essential semantic instability of language, or the relativity of meaning to context, undermines even the assumption that identical meanings are exchanged in communication.

In reply to this, Habermas suggests that joking, playful, ironic and fictional uses of language depend on an intentional confusion of contrasting modes correlated to validity-claims (being/illusion, is/ought, essence/appearance), which is simultaneously seen through as such. These uses therefore depend on a prior recognition of the distinctions between validity dimensions. Habermas also emphasizes that belief in the identity of communicated meanings has no *metaphysical* status, but is merely a pragmatic presupposition which can (indeed must) be unproblematically relied on, up to the point where communication actually breaks down.

3 Apel's approach and his critique of Habermas

Habermas' account of communicative rationality has developed over the years in tandem with that of his Frankfurt colleague Karl-Otto Apel. However, Apel's programme is more strongly connected to the tradition of transcendental philosophy (Apel prefers the term 'transcendental pragmatics' to Habermas' 'universal pragmatics' for the general theory of communicative competence). The key conflict between Apel and Habermas concerns their attitude to the possibility of an 'ultimate grounding' (*Letztbegründung*) for a universalistic moral standpoint. In a critical essay (1992), Apel contends that Habermas is mistaken to play down the 'transcendental difference' between empirically testable reconstructions of communicative competence and philosophical reflection which reveals the unconditional validity of certain basic presuppositions of argumentation. Apel sees here a confusion which leads to paradoxical consequences - for example, that the 'principle of fallibilism' (the principle that no empirical consensus can be immune to revision) must itself be regarded as fallible. Habermas, however, replies that the normative preconditions of communication should be considered only as 'factically indispensable': they cannot be imagined as inoperative given our human socio-cultural form of life. To show this, however, we do not need to set philosophical truths in a position of qualitative precedence over knowledge obtained by scientific methods.

The concept of communicative rationality has thus been, and continues to be, the subject of intense debate. These reflect that fact that the project of defining a concept of rationality more encompassing that instrumental rationality, one internally connected to the idea of *reciprocal* and *egalitarian* relations between human beings, has deep roots in the German philosophical tradition since Kant. Because of this, whatever future transformations this project may undergo, it is unlikely to be entirely abandoned.

See also: Habermas, J.; Rationality: belief; Reasoning/rationality: practical

PETER DEWS

References and further reading

Adorno, T.W. and Horkheimer, M. (1944) *Dialectic of Enlightenment*, London: Verso, 1979.(Referred to in §1. Presents the classic Frankfurt School critique of instrumental reason.)

Apel, K.-O. (1973) *Transformation der Philosophie*, Frankfurt am Main: Suhrkamp Verlag, 2 vols; partial trans. G. Adey and D. Frisby, *Towards a Transformation of Philosophy*, London: Routledge & Kegan Paul, 1980. (Sets out Apel's theory of communicative rationality, based on the regulative idea of an unlimited communication community.)

Apel, K.-O. (1992) 'Normative Begründung der "Kritischen Theorie" durch Rekurs auf Lebensweltliche Sittlichkeit? Ein Transzendental-pragmatisch orientierter Versuch mit Habermas gegen Habermas zu denken' ('Normatively Grounding "Critical Theory" through Recourse to the Lifeworld? A Transcendental-Pragmatic

Attempt to Think with Habermas against Habermas'), in A. Honneth, T. McCarthy, C. Offe and A. Wellmar (eds) *Zwischenbetrachtungen. Im Prozess der Aufklärung (Philosophical Interventions in the Unfinished Project of Enlightenment)*, Frankfurt am Main: Suhrkamp Verlag; trans. B. Rehg, Cambridge MA: MIT Press, 125-72.(Referred to in §3. The fullest statement of Apel's critique of Habermas.)

Cooke, M. (1994) *Language and Reason: A Study of Habermas' Pragmatics*, Cambridge MA: MIT Press.(A detailed, sympathetic but critical analysis of Habermas' concept of communicative rationality.)

Habermas, J. (1968) *Erkenntnis und Interesse*, Frankfurt am Main: Suhrkamp Verlag; trans. J.J. Shapiro as *Knowledge and Human Interests*, London: Heinemann, 1971.(Referred to in §1. Develops the distinction between instrumental action and communicative action as two distinct forms of action.)

Habermas, J. (1976) 'Was heißt Universalpragmatik', in K.-O. Apel (ed.) *Sprachpragmatik und Philosophie*, Frankfurt am Main: Suhrkamp Verlag; trans. T. McCarthy as 'What is Universal Pragmatics?', in *Communication and the Evolution of Society*, London: Heinemann, 1979, 1- 68.(Habermas' most detailed statement of his programme for a universal pragmatics)

Habermas, J. (1981) *Theorie des Kommunikative Handelns. Band 1: Handlungsrationalität und gesellschaftliche Rationalisierung*, Frankfurt am Main: Suhrkamp Verlag; trans. T. McCarthy as *Reason and the Rationalization of Society*, vol. 1 of *The Theory of Communicative Action*, Boston MA: Beacon Press, 1984.(Contains important discussions of the concept of communicative rationality.)

Habermas, J. (1988) 'Handlungen, Sprechakte, sprachlich vermittelte Interaktionen und Lebenswelt', in *Nachmetaphysiches Denken (Postmetaphysical thinking)*, Frankfurt am Main: Suhrkamp Verlag. (Source of the second quotation in §2 above.)

Schnädelbach, H. (1991) 'Transformation der Kritischen Theorie' ('The Transformation of Critical Theory'), in A. Honneth and H. Joas (eds) *Kommunikatives Handeln (Communicative Action)*, trans. J. Gaines and D.L. Jones, Cambridge: Polity Press, 7-22.(Referred to in §2. Criticizes Habermas for suggesting that identifying grounds involves evaluating them, thereby making any description of communicative rationality normative.)

Wellmer, A. (1986) *Ethik und Dialog. Elemente des moralischen Urteils bei Kant und in der Diskursethik*, Frankfurt am Main: Suhrkamp Verlag; trans. D. Midgeley as 'Ethics and Dialogue: Elements of Moral Judgement in Kant and Discourse Ethics', ch. 4 of *The Persistence of Modernity*, Cambridge: Polity Press, 1991, 113-231.(Referred to in §2. An important critique of the idealizations built into Habermas' theory of communicative rationality.)

Compositionality

A language is compositional if the meaning of each of its complex expressions (for example, 'black dog') is determined entirely by the meanings its parts ('black', 'dog') and its syntax. Principles of compositionality provide precise statements of this idea. A compositional semantics for a language is a (finite) theory which explains how semantically important properties such as truth-conditions are determined by the meanings of parts and syntax. Supposing English to have a compositional semantics helps explain how finite creatures like ourselves have the ability to understand English's infinitely many sentences. Whether human languages are in fact compositional, however, is quite controversial.

It is often supposed that meaning, within a context of use, determines truth-value. Meaning and context are also generally thought to determine what terms refer to, and what predicates are true of - that is, to determine extensions. Because of this, compositionality is generally taken to license a principle of substitutivity: if expressions have the same meaning, and substituting one for another in a sentence does not change the sentence's syntax, the substitution can have no effect on truth. Given this, the idea that meaning is compositionally determined constrains what can be identified with meaning. For example, expressions with the same extension cannot always be substituted for one another without change of truth-value (see Propositional attitude statements §2; Modal logic, philosophical issues in §3).

Natural languages obey a principle of compositionality (PC) only if something more 'fine-grained' than extensions plays the role of meanings, so that expressions with the same extension can still have different meanings. Some say that no matter how fine-grained we make meaning, there will be counterexamples to the claim that meaning is compositionally determined. For example, 'woodchuck' and 'groundhog' are synonymous. But the sentences 'Mae thinks Woody is a woodchuck' and 'Mae thinks Woody is a groundhog' apparently may diverge in truth-value. (Certainly Mae may assert, with understanding, that Woody is a woodchuck, while denying, with understanding, that Woody is a groundhog. For she can understand 'woodchuck' and 'groundhog' but not realize that they are synonymous.) Such arguments lead to the conclusion that meaning is not compositionally determined (or that different forms cannot share meanings).

There is no consensus about such arguments. All agree PCs need *some* restrictions - they do not, for instance, apply to words in quotational constructions. So if 'thinks' is implicitly quotational, the argument loses its force (see Use/mention distinction and quotation). If 'thinks', as Frege held, produces a (more or less) systematic shift in the semantic properties of the expressions in its scope, this perhaps reduces the threat to compositionality (see Sense and reference §§4, 5). Some find such accounts of 'thinks' incredible (see Davidson 1984).

Linguists often understand the claim that a language is compositional as asserting an extremely tight correspondence between its syntax and semantics (see Syntax). A (simplified) version of such a claim is that (after disambiguating simple word forms), there is, for each (simple) word, a meaning and, for each syntactic rule used in sentence construction, an operation on meanings, such that the meaning of any sentence is mechanically determined by applying the operations on meanings (given by the rules used in constructing the sentence) to the meanings of the simple parts. (Often a host of extra restrictions are incorporated. For example: the operations may be limited to applying function to argument; the order in which operations are applied may be settled by the structure of the sentence.) Some see such principles as providing significant constraints on semantical theories, constraints which may help us decide between theories which are in other respects equivalent (for example, Montague 1974).

Such strong PCs imply that every ambiguous sentence is either syntactically ambiguous or contains simple expressions which are themselves ambiguous (see Ambiguity). There are putative counterexamples to this. For instance, it has been alleged that 'The women lifted the piano' has two meanings, a 'collective' one (the group lifted it) and a 'distributive' one (each individual lifted it) (see Pelletier 1994).

Context sensitivity makes the formulation of PCs a delicate matter. In some important sense, differing uses of 'that is dead' have like meanings and syntax. But obviously they may differ in truth; presumably such differences may arise even within one context. How can this be reconciled with compositionality?

We might (implausibly) say that demonstrative word forms, strictly speaking, lack meanings (only tokens have such). An alternative (see Kaplan 1989) sees the meaning of the word type 'that' as 'incomplete', needing contextual supplementation to determine a meaning of the sort other terms have. A third strategy insists that if 'that is dead but that is not' is true, its 'that's are different demonstrative word types. (Advocates of this view are good at seeing very small subscripts.) This view makes it generally opaque to speakers whether tokens are tokens of the same word type, and thus opaque whether their arguments are logically valid. A fourth strategy sees context as providing an assignment of referents to *uses* of a type in a context. On most ways of working this out, it seems there is no true fleshing out of the principle: the meaning of a sentence type (a) is determined by the meanings of parts and syntax, and (b) determines, in context, the sentence's truth-value. This approach also seems to undermine the idea that an argument like *that is dead; so that is dead* is formally valid (see Demonstratives and indexicals).

Even if relatively tight PCs should prove untenable (as claims about natural languages), it would not follow that natural languages did not have compositional semantics - finite theories which assign truth-conditions and other semantically relevant properties to sentence types or their uses. It has been argued that only if they do would it be possible for finite creatures like ourselves to learn them (since the languages involve infinite pairings of sounds and meanings); analogous arguments hold that our ability to understand (in principle) natural languages requires the existence of such semantics. Such arguments apparently presuppose that learning and understanding a language requires knowing a theory from which semantic facts (such as the fact that, for any speaker x and time t, a use of 'J'existe encore' by x at t says that, at t, x still exists) are deducible. While such a view is not implausible, alternative plausible views of competence do not support this sort of argument. For example, one might identify linguistic competence with the possession of *syntactic* knowledge (or even just with the possession of syntactic abilities) by someone with appropriate social and environmental relations and behavioural dispositions. Whether natural languages have compositional semantics, and whether the meanings of their sentences are determined simply by the meanings of their parts and syntax, is still not settled.

See also: Meaning and truth; Semantics

<div align="right">MARK RICHARD</div>

References and further reading

Davidson, D. (1984) *Inquiries into Truth and Understanding*, Oxford: Oxford University Press.(Essay 1 argues that learnable languages have compositional semantics; essay 7 argues against Fregean accounts of 'thinks'.)

Grandy, R. (1990) 'Understanding and the Principle of Compositionality', in J.E. Tomberlin (ed.) *Philosophical Perspectives*, Atascadero, CA: Ridgeview Publishing Company, vol. 4, 557-72.(Defends compositional semantics against the attacks of Schiffer and others.)

Kaplan, D. (1989) 'Demonstratives', in J. Almog, J. Perry and H. Wettstein (eds) *Themes From Kaplan*, Oxford: Oxford University Press.(Gives Kaplan's views of demonstrative meaning.)

Montague, R.M. (1974) *Formal Philosophy*, New Haven, CT: Yale University Press. (Essays 6-8 give formulations of compositionality principles. Very difficult.)

Partee, B. (1984) 'Compositionality', in F. Landman and F. Veltman (eds) *Varieties of Formal Semantics*, Dordrecht: Foris, 281-311.(Evaluates Montague-style principles of compositionality. Difficult but illuminating.)

Pelletier, F.J. (1994) 'The Principle of Semantic Compositionality', *Topoi* 13: 11-24. (A good introduction to arguments for and against principles of compositionality.)

Schiffer, S. (1987) *Remnants of Meaning*, Cambridge, MA: MIT Press.(Contains arguments against the existence of compositional semantics for natural languages.)

Counterfactual conditionals

'If bats were deaf, they would hunt during the day.' What you have just read is called a 'counterfactual' conditional; it is an 'If…then…' statement the components of which are 'counter to fact', in this case counter to the fact that bats hear well and sleep during the day. Among the analyses proposed for such statements, two have been especially prominent. According to the first, a counterfactual asserts that there is a sound argument from the antecedent ('bats are deaf') to the consequent ('bats hunt during the day'). The argument uses certain implicit background conditions and laws of nature as additional premises. A variant of this analysis says that a counterfactual is itself a condensed version of such an argument. The analysis is called 'metalinguistic' because of its reference to linguistic items such as premises and arguments. The second analysis refers instead to possible worlds. (One may think of possible worlds as ways things might have gone.) This analysis says that the example is true just in case bats hunt during the day in the closest possible world(s) where they are deaf

1 Classification issues and familiar logical connectives

'Counterfactual' conditionals serve to summarize hypothetical reasoning and to justify or explain action and inaction ('I didn't tell you because, had I told you, everyone in the office would have known'). They serve to state particular causal contingencies ('If the battery were flat, the starter would not turn') and various kinds of necessity ('If he were dead, she would be a widow'). Counterfactuals have also been enlisted for philosophical analyses of dispositions ('*X* is soluble just in case it would dissolve if it were immersed in water'), causality ('Had *A* not occurred, *B* would not have occurred'), freedom of action ('I would have acted otherwise if I had so chosen'), knowledge ('If *p* were false, Mary would not believe that *p*') and laws of nature, which have been characterized as generalizations that are not merely true but support counterfactuals.

Counterfactuals are commonly contrasted with indicative conditionals (see Indicative conditionals). The following pair conveys a feeling for the difference:

(1) If Shakespeare hadn't written *Hamlet*, then someone else would have (counterfactual).
(2) If Shakespeare didn't write *Hamlet*, then someone else did (indicative).

Instead of 'counterfactual conditional', sometimes the term 'subjunctive conditional' is used. The former refers to a property of the statement made by uttering a conditional sentence in an appropriate context, the latter refers to the grammatical form of the sentence uttered. Neither label is satisfactory. A problem with 'counterfactual' is that not all assertions of counterfactuals are, or are meant to be, counter to fact, as, for example, 'If you did the cooking, I would clean up afterwards' (put forward as a suggestion before the cooking has started). 'Subjunctive' is problematic because the grammatical form in question is not the subjunctive in English (where this form is all but extinct) nor in other European languages (where it is widely used). Moreover, 'subjunctive' contrasts with 'indicative', yet indicatives such as (3) below go much more naturally with 'subjunctives' such as (1) than with indicatives such as (2):

(3) If Shakespeare does not write *Hamlet*, then someone else will.

It is surprisingly difficult to characterize the difference between the two kinds of conditional. It is not even agreed at which linguistic level - morphosyntactic, semantic or pragmatic - the distinction is to be drawn. We will content ourselves here with defining counterfactual *sentences* as the likes of (1) above. Our question is then which *statements* these sentences normally serve to express.

In the view of most authors, an analysis must state non-circular and informative truth-conditions. It should also be 'compositional': the truth-conditions of ' $A \rightarrow B$' should be a function of the truth-conditions of *A* and *B* (see Compositionality). (Read ' $A \rightarrow B$' as 'If it were the case that *A*, then it would be the case that *B*'.) The material conditional of classical logic and the strict conditional of modal logic both meet these demands, but their truth-conditions are unsuitable. A material conditional is true just in case either its antecedent is false or its consequent true. Since most counterfactuals have false antecedents, they would come out true irrespective of their consequents if they were treated as material conditionals. But many are not true (for example, (1) above); hence the material conditional cannot be the right analysis. A strict conditional is true just in case the corresponding

material conditional is necessarily true, or, in possible worlds parlance, just in case in all worlds in which the antecedent is true the consequent is also true. If the material conditional is too easy to satisfy, the strict conditional is too hard. Consider

(4) If fewer people smoked, fewer would contract lung cancer,

which is true if any counterfactual is. Yet there are possible worlds where the antecedent is true and the consequent false, for example, worlds in which lung cancer is caused by eating chocolate. So the strict conditional is too strict. We might confine our attention to less exotic worlds; ones governed by the same natural laws as the actual world. But that still would not do because among the lawful worlds where fewer people smoke there are worlds where more people work in coal mines and the overall incidence of lung cancer is the same. The worlds in which we are interested must resemble the actual world more closely. We imagine that fewer people smoke in these worlds, but that otherwise things are very much as in the actual world. It is this idea of a minimal departure from actuality (or purported actuality) that underlies, in one form or another, all theories of counterfactuals.

2 The metalinguistic view

Nelson Goodman (1954) put forward an influential view in which ' $A \rightarrow B$' is true if and only if (1) there is an appropriate subset of all the true initial conditions that, in conjunction with A and the laws of nature, entails B; and (2) there is no other subset of the initial conditions that in the same manner entails not-B. The minimal departure from actuality lies in the appropriate choice of initial conditions. Two problems beset Goodman's proposal: the characterization of the relevant initial conditions and the definition of laws. The second problem is serious, but not as serious as the first.

Apart from being true, the relevant initial conditions must be jointly compatible with A (as long as A is consistent), for otherwise any consequent whatever could be inferred and the corresponding counterfactual would be true. The conditions must also be compatible with not-B, for otherwise A and the laws would play no role in the inference to B - B would follow from the initial conditions by themselves. ('Compatible' statements are ones which do not jointly entail a contradiction.) But compatibility alone is not a strong enough condition. Consider

(5) If match m had been struck, it would have lighted.

Let O = 'oxygen is present', D = 'm is dry', S = 'm is struck', L = 'm lights' and N = 'If O, D and S, then L' (our natural law). Suppose that we start with the situation of O, D and N being true and that the match has not been struck and has not lighted. O, D and N are jointly compatible with S and jointly compatible with not-L, and together with S they imply L. Therefore (5) should be true.

However, not-L is also true, not-L, O and N are jointly compatible with S and jointly compatible with D; and O, N and not-L together with S imply not-D. Thus

(6) If match m had been struck, it would not have been dry

should be true as well. This is unacceptable. Hence logical compatibility is not enough. The relevant conditions must be jointly tenable or 'cotenable' with the antecedent. (B is said to be cotenable with A if it is not the case that B would be false if A were true.) In this case, not lighting is compatible with, but not cotenable with, striking, because, had m been struck, it would have lighted; so (5) is true and (6) is not. But now we are caught in a circle or regress because cotenability is defined in terms of counterfactuals, while counterfactuals are defined in terms of cotenability. Goodman saw no way out of this quandary.

3 Possible worlds semantics

In the 1960s, modal logicians turned talk of possible worlds into a powerful analytic tool. Robert Stalnaker and David Lewis applied it to the analysis of counterfactuals. Both saw counterfactuals as devices to locate the actual world within a certain similarity structure in logical space. The intuition they built on is that ' $A \rightarrow B$' is true in the actual world just in case B is true in the most similar A-world(s) (an A-world being a world where A is true). Stalnaker (1968) represents the minimal departure from actuality by a selection function on possible worlds, Lewis (1973) by ordering worlds so that given any two worlds u and v, either u is more similar to the actual world than v, or vice versa, or else u and v are equally similar to the actual world. For Stalnaker, ' $A \rightarrow B$' is true if and only if

B is true at $f(A, w^*)$, the world picked out by applying the selection function *f* to antecedent *A* and the actual world w^*. *f* selects a world in which *A* is true and which, intuitively, resembles w^* as closely as the truth of *A* permits. For Lewis, ' $A \to B$ ' is true if and only if either there is no *A*-world, or some *A* & *B*-world is more similar to the actual world than any *A* & not-*B*-world. Lewis' analysis makes ' $A \to B$ ' trivially true when *A* is impossible, which is when there is no *A*-world. So does Stalnaker's analysis because, for impossible *A*, *f* selects 'the impossible world' in which every statement is true.

An argument form is valid if and only if it has no instance with true premise(s) and a false conclusion. The following is an example of an invalid argument:

If Ronald Reagan had been born a Russian, he would have been a communist.
If he had been a communist, he would have been a traitor.
Therefore, if he had been born a Russian, he would have been a traitor.

The premises seem true, but not the conclusion. The argument is an instance of 'hypothetical syllogism':

$$A \to B, \quad B \to C \quad \therefore \ A \to C$$

The possible worlds theory explains how hypothetical syllogism can fail to be valid. World *c*, the closest world where Reagan was a communist, is closer to the actual world than world *r*, the closest world where he was born a Russian. In *c*, his career is as we know it, except for that traitorous affiliation. In *r*, he grows up in Russia, becomes a loyal party member there and never dreams of being president of the USA. So *c* makes the second premise of the argument true, while *r* makes the first premise true and the conclusion false.

On the Lewis-Stalnaker analyses, a number of argument forms that are valid for the material and various strict conditionals are invalid. Hypothetical syllogism is only one of them. Others are:

$A \to B \ \therefore \ \text{not-}B \to \text{not-}A$ ('contraposition')
$A \to B \ \therefore \ C \to B$, where C entails A ('strengthening the antecedent')
$A \to (B \to C) \ \therefore \ (A \ \& \ B) \to C$ and
$(A \ \& \ B) \to C \ \therefore \ A \to (B \to C)$ ('import-export')

For practical purposes, the invalid argument forms can often be replaced by similar valid ones, for example, hypothetical syllogism by

$$A \to B, \quad (A \ \& \ B) \to C \quad \therefore \ A \to C.$$

The distinction between valid and invalid argument forms ought to match natural language usage. Examples like the one about Reagan therefore carry heavy weight. If they were rejected and any of the argument forms in question added to the logics that emerge from the Lewis-Stalnaker analyses, ' $\to$ ' would collapse into the material conditional. Attacks on the counterexamples are therefore indirect attacks on the core of the analyses. (For more discussion, see Stalnaker 1984.)

4 Similarity

Lewis allows ties in the relative similarity between worlds, as well as no limit to how closely worlds can resemble each other. Stalnaker's theory can be seen as a special case of Lewis' that does not make these allowances. For Stalnaker, there is to be always exactly one most similar antecedent-world. The difference is reflected in a disagreement about the principle of 'conditional excluded middle' which says that ' $A \to B$ ' and ' $A \to \text{not-}B$ ' cannot both be false. The principle holds in Stalnaker's view but not in Lewis'. In Lewis' view, there may be, for any *A* & *B*-world, an equidistant or closer *A* & not-*B*-world and vice versa, so that both counterfactuals can be false. For Stalnaker, there is exactly one closest *A*-world in which either *B* or not-*B* is true. Consider the pair

If Bizet and Verdi had been compatriots, Bizet would have been Italian.
If Bizet and Verdi had been compatriots, Verdi would have been French.

We hesitate to affirm either. Doesn't this show that Lewis is right? Not necessarily. We also hesitate to deny either of the two, and the best explanation may be that the relevant standard of similarity between worlds has not been

made clear, so that we do not know exactly what claims are at issue. Conditional excluded middle holds relative to a given selection function *f*, but it does not apply to cases where *f* has not been determined. Any theory of counterfactuals must allow for a good measure of indeterminacy if it is to account for actual usage.

Goodman's theory can be transposed into the possible worlds framework. With the right definition of cotenability, it becomes equivalent to Lewis'. Does Lewis therefore solve Goodman's problem about cotenability? In some sense, yes, because there is no longer a circle in the truth-conditional schema. In another sense, no, because Goodman's worry was less about the abstract form of the analysis than about how particular cotenability judgments can be grounded. For an answer to this question, the notion of relative similarity between worlds is probably too vague.

The relevant notion of similarity cannot be the one we use when we compare ordinary objects. For otherwise we should reject

If Nixon had pushed the button, there would have been a nuclear holocaust,

because among the worlds where Nixon pushes the ominous button those where nothing happens are more similar to ours, in the everyday sense of similarity, than those where the doomsday machinery goes off. But given that there is (or was) such a button, the counterfactual looks true. Lewis suggests that the similarity relation relevant for counterfactuals has the most similar worlds agree with the actual world throughout the past until shortly before the time of the antecedent. Then a small miracle (a deviation from the laws of the actual world, not from those of the world in question) brings about the antecedent, and from then on history evolves in accordance with the laws of the actual world. The proposal has limited scope. It does not apply to 'counterlegals' about what would be the case if this or that law of nature were different. It also does not deal with disjunctive antecedents pertaining to different times. Maximizing agreement throughout the past would naturally lead here to a selection of worlds in which only the disjunct pertaining to the latest time holds. Other authors favour total agreement in laws over miracles of any size. Miracles prevent a counterfactual extrapolation of what the past would have had to have been in order to bring about the antecedent. Yet such 'backtracking' is perfectly intelligible. Jim and Jack quarrelled yesterday and Jack is still furious. If Jim asked a favour of Jack today, Jack would oblige him none the less. For in order that Jim ask, there would have had to have been no quarrel before. The significance of backtracking and Lewis' proposal about similarity remain contested.

5 Comparisons

Metaphysical and other qualms about the possible worlds analyses rekindled an interest in metalinguistic theories in the 1980s. There have also been proposals for an epistemic analysis based on the idea that ' $A \rightarrow B$' is 'acceptable' with respect to a body of beliefs *K* just in case *K*, minimally changed so as to accommodate *A*, entails *B*. Formally, most of these theories are equivalent to Lewis' or slight variants thereof, but they avoid the ontological cost of possible worlds. In so far as they do not provide truth-conditions, the epistemic theories have difficulty explaining the role of counterfactuals embedded in truth-conditional constructions such as conjunctions or disjunctions. Moreover, the idea of revising a body of beliefs seems better suited for evaluating indicative conditionals than counterfactuals. For example, if you add to your present beliefs the assumption that Shakespeare did not write *Hamlet* and make the minimal changes necessary to restore consistency, you will infer that someone else must have written the piece because you have no reason to give up your belief in its existence. This yields the correct judgment for the indicative (2) but not for the counterfactual (1). It is not obvious how to adapt procedures of belief change to counterfactuals.

Goodman and many others believed that an analysis of counterfactuals was indispensable for the philosophy of science. Few still share this opinion. The cotenability problem, the indeterminacy of many counterfactuals, and doubts about possible worlds and their similarities have convinced many that counterfactuals do not belong in the ultimate scientific description of the world. In decision and game theory, by contrast, the peculiarity and importance of counterfactual reasoning has only just begun to be appreciated. The solution of decision problems requires answers to questions about situations that, as a result of the very decision at issue, will be counterfactual. Counterfactuals also continue to play an important role in philosophical analyses, and their use and misuse in daily commerce is largely unhampered by difficulties of interpretation.

See also: Logical and mathematical terms, glossary of; Possible worlds; Relevance logic and entailment

FRANK DÖRING

References and further reading

Edgington, D. (1995) 'On Conditionals', *Mind* 104.414: 235-329.(Gives an excellent overview and synthesis of the various issues surrounding conditionals.)

Goodman, N. (1954) *Fact, Fiction, and Forecast*, Atlantic Highlands, NJ: Athlone Press.(See §3. A non-technical classic.)

Harper, W.L., Stalnaker, R. and Pearce, G. (eds) (1981) *IFS: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht and Boston, MA: Reidel.(A collection of important papers of varying degrees of technicality.)

Jackson, F. (ed.) (1991) *Conditionals: Oxford Readings in Philosophy*, Oxford: Oxford University Press.(A collection of important papers, including Lewis' on the similarity between worlds.)

Lewis, D.K. (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press. (See §4. The philosophical discussion demands no technical expertise, but the details of the semantics and the logic get very technical.)

Sanford, D.H. (1989) *If P then Q: Conditionals and the Foundations of Reasoning*, London and New York: Routledge.(An informative historical overview.)

Stalnaker, R. (1968) 'A Theory of Conditionals',*Studies in Logical Theory: American Philosophical Quarterly Monograph Series* 2: 98-112; repr. in W.L. Harper, R. Stalnaker and G. Pearce (eds) *IFS: Conditionals, Belief, Decision, Chance, and Time*, Dordrecht and Boston, MA: Reidel, 1981, 41-56.(See §4. The first statement of Stalnaker's position; no technical expertise required.)

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: Bradford Books.(An excellent non-technical discussion of counterfactual and indicative conditionals and their role in reasoning.)

Criteria

The concept of criteria has been interpreted as the central notion in the later Wittgenstein's account of how language functions, in contrast to the realist semantics of the Tractatus. According to this later account, a concept possesses a sense in so far as there are conditions that constitute non-inductive evidence for its application in a particular case. This condition on a concept's possessing a sense has been thought to enable Wittgenstein to refute both solipsism and scepticism about other minds. There are powerful objections to this conception of criteria, which have led some philosophers to look for an alternative account of the role of criteria in Wittgenstein's later philosophy.

1 Criteria and scepticism

Norman Malcolm (1963) sets out the central argument by means of which Wittgenstein allegedly refutes solipsism and scepticism about other minds. Wittgenstein is held to expose the incoherence of the argument from analogy, and thereby to show that if we make the assumption that we know, for example, what pain is only from our own case, then we could never transfer the idea to others. This allegedly forces the defender of the view that 'pain' refers to a private object within a psychological realm into endorsing solipsism. Wittgenstein is then held to use the private language argument to show that solipsism is incoherent (see Wittgenstein, L. §13; Private language argument).

The solipsist has a concept of pain only if it makes sense to talk of identifying a sensation as a sensation of pain correctly or incorrectly. But what does it mean to say that the solipsist has correctly identified a given sensation as one of pain? The solipsist has neither a standard, a sample nor a customary practice of using the word 'pain' against which the inclination to apply the concept can be judged. There is nothing independent of this inclination that enables the solipsist to determine whether the application is correct or incorrect. But in so far as there is nothing to determine whether the application is correct or incorrect, the idea of using the word correctly or incorrectly makes no sense. And in so far as the solipsist lacks a criterion of correct application, the solipsist lacks a concept of pain.

Thus, the assumption that one knows what pain is only from one's own case is shown to lead to a contradiction: if we know about sensations only from our own case, then we cannot know about them even in our own case. It follows, Malcolm argues, that it is a condition of our being able to attribute sensations to ourselves that there should exist *criteria* on the basis of which we are necessarily justified in ascribing them to others. When these criteria are fulfilled, speakers do not merely have good inductive evidence that another is in pain (as the argument from analogy holds); they possess evidence that *establishes beyond question* that another is in pain.

This initially powerful rebuttal of scepticism is weakened, however, when Malcolm is forced to concede that there is a wide range of circumstances (being in a play, being hypnotized, pretending and so on) in which behaviour that would, in other circumstances, constitute criterial evidence for another's being in pain, does not count as fulfilling the criteria. Thus, what looks initially like an argument that establishes that there must be logically necessary and sufficient conditions for another's being in pain - that is, that there must be circumstances in which the behavioural evidence would make it a contradiction for the sceptic to deny that another is in pain - receives a fatal qualification. For given that a doubt may arise over whether the circumstances are ones in which criteria are satisfied (or really satisfied), it seems that the sceptic could always raise a question about any particular case. What looked like an argument to establish that our belief in another's pain is, on particular occasions, logically justified, is now reduced to the much weaker claim that there must be circumstances in which a speaker *accepts* that the criteria for another's being in pain are fulfilled.

There are a number of familiar objections to the argument that Malcolm attributes to Wittgenstein. First of all, it has been seen as amounting to little more than a version of verificationism, and as such it is thought not to warrant the status or the degree of significance that Malcolm ascribes to it. Secondly, it has been objected that the promise to provide a refutation of scepticism is undermined by the qualifications set out in the previous paragraph. It looked as if criteria were to enable us to show that our belief in another's pain is logically justified, but it turns out that the satisfaction of criteria on particular occasions is something that we must ultimately just accept, even though a doubt is, in some sense, still possible. The sceptic could be forgiven for regarding this as vindication,

rather than refutation, for there remains a gap between evidence and truth that must be filled by brute faith that our criteria are satisfied. At this point we seem justified in returning to Wittgenstein's texts to see if there is an alternative interpretation of his use of the concept of criteria.

2 Criteria and grammar

The principal concerns of Wittgenstein's later philosophy are the false pictures of language and of psychological phenomena that plague us in philosophy. These myths and false pictures arise, Wittgenstein believes, when we treat the phenomena that characterize our human form of life - language, consciousness, thought, action and so on - as things whose essence lies hidden from us. We both discover the emptiness of the pictures we construct, and achieve the understanding that we seek, through what Wittgenstein calls a 'grammatical investigation'. His idea of a grammatical investigation is characteristically rich. He describes such an investigation as one in which 'we remind ourselves of the kind of statement that we make about phenomena'. Because the essence of language and psychological phenomena lies open to view in the forms of our ordinary practice, it is through a careful description of our actual employment of expressions that we come to understand the nature of these phenomena. The understanding which a grammatical investigation achieves simply sets before us 'what already lies open to view', but in such a way that it no longer puzzles us or cries out for further explanation. Part of the difficulty in understanding Wittgenstein's later philosophy is that he does not believe that the understanding which a grammatical investigation achieves can be expressed in the form of a clear, unambiguous description of the structure and function of language in general, or of our psychological language game in particular.

It is possible to see Wittgenstein's concept of criteria not as a theoretical term of art but as linked in important ways with this idea of a grammatical investigation. Thus, asking for the criteria that ordinarily establish, in the sense of *identify*, something as an instance of a specific kind of thing (as a mistake, as expecting someone between 4.00 and 4.30, as understanding the order 'Add two' in the way that it was meant and so on) is one particular form of grammatical investigation. Other forms include asking how we teach someone a concept, asking how we would explain it, asking whether we would apply it in certain non-standard cases, asking whether certain facts being different would make it unusable, comparing it with a concept that Wittgenstein invents, and so on. Wittgenstein uses these forms of investigation to induce the clarified vision of the workings of our ordinary language games that he believes is essential to our coming to recognize that everything we need to understand language and psychological phenomena is already there before our eyes.

One of the principal lessons of the grammatical investigation of our ordinary criteria is that the criteria governing our concepts are much more complicated, our language games much more subtle and involved, than at first appears. In the case of psychological concepts this complexity is expressed, not only by the first-person-third-person asymmetry which characterizes these concepts (for example in the case of pain, I do not attribute pain to myself on the basis of observation, the concept of a mistake does not apply in the first-person case, and it makes no sense for me to doubt whether I am in pain), but also by an indeterminacy and uncertainty which Wittgenstein again takes to be a defining feature of our psychological language game. Wittgenstein believes that it is in part these grammatical features of our psychological concepts that prompts the philosopher to form a picture of the mental as an inner realm. For not only is the first-person-third-person asymmetry captured in the distinction between inner and outer, but the idea of the inner allows us to interpret the indeterminacy and uncertainty that is inherent in the language game as merely epistemic: the facts are there alright, but hidden beneath the surface that the other presents to the world.

Wittgenstein's grammatical investigation is directed at a complete overcoming of the philosophical myth of the inner. On the one hand, he tries to show that the picture of a determinate inner realm not only plays no role in our ordinary language game, but that it has no real application. His tactic here is not to try to refute the philosopher's claims, but to develop and explore them in such a way that we come to see, not only that no such thing is ever ordinarily in question, but that the picture of a private inner realm, which initially seems so clear to us, has no real content. He does not argue that no precise system of concepts linked, for example, with exact physiological readings of what is occurring in our brain and nervous system could never be constructed, but he tries to show us that such a system of concepts would not tell us what pain, or thought, or understanding really are; it would in no sense be a refinement of our system of psychological concepts, or constitute an explication of the phenomena that make up our form of life.

On the other hand, Wittgenstein's detailed exploration of the complex, infinitely nuanced language game that the process of acculturation initiates us into, gradually enables us to see how our language game actually functions, and to resist the temptation to misrepresent it in the myth of the inner. He works to overcome our sense that to abandon the idea of a private inner realm is to deny something vital, something without which we are mere machines. He uses the techniques of grammatical investigation to show that our philosophical picture of the inner is grounded in a fundamentally mistaken idea of the relation between our ordinary psychological concepts and the characteristic forms of movement, gesture and expression of the living bodies of humans and other animals. The division which we are mistakenly inclined to draw between an inner and outer realm is shown to be one that is actually grounded in the grammatical distinctions between psychological and non-psychological concepts.

Thus, we gradually come to see that our ordinary psychological language game is distinctive, not in that it describes a hidden realm of facts, but in its grammar. The grammar of our psychological language game reveals the nature of the facts it describes, and in recognizing the distinctiveness of this grammar, we recognize the distinctive form of the phenomena with which we are concerned. Thus, the first-person-third-person asymmetry which characterizes psychological concepts is not something that needs to be explained by (the empty) appeal to a special sort of fact (private facts); rather, this asymmetry itself reveals the distinctive form of mental phenomena.

Likewise, pretence, deceit, betrayal, as well as the ordinary dissemblings that constitute polite behaviour, are part of the form of our psychological language game, and not an unfortunate or incidental addition to it. These phenomena should not lead us to downgrade the 'outer evidence', but to recognize the subtlety and the complexity of the criteria we operate with. Sometimes we are sure of our judgment that another's feeling is genuine, sometimes anyone who is not a lunatic will share our certainty, but at other times we are uncertain and the judgment of different speakers may vary. What Wittgenstein tries to show is that this reflects the essential complexity and ambiguity of our criteria, and not the hiddenness of the facts; our uncertainty arises from the complexity and the subtlety of our relations with others and of the patterns that our psychological concepts require us to discern, and not from the indirectness of our evidence.

Wittgenstein's later philosophy of psychology clearly does constitute a concerted attack on the myth of the inner. This attack does not, however, take the form of a theory of how language functions that imposes on psychological concepts the requirement that what is inner is criterially linked with what is outer. The understanding of psychological phenomena that Wittgenstein's grammatical investigation offers is, in a sense, much more radical than this suggests. What we are gradually brought to see is that the nature of psychological phenomena is not hidden; that the evidence for their existence is not indirect, but involved and complex; that everything that we need to understand the essence of these phenomena is already there in the distinctive grammar of our language game, as this is revealed, for example, by the nature of the criteria with which we operate.

See also: Contextualism, epistemological; Other minds

MARIE McGINN

References and further reading

Albritton, R. (1959) 'On Wittgenstein's Use of the Term "Criterion"', *The Journal of Philosophy* 56: 845-57. (Explores the view that criteria are logically necessary and sufficient conditions.)

Cavell, S. (1979) *The Claim of Reason*, Oxford: Clarendon Press.(This provides the excellent criticisms of Malcolm and the best introduction to the anti-Malcolm view of criteria. See especially part 1, chapters 1 and 2.)

Malcolm, N. (1963) 'Wittgenstein's *Philosophical Investigations'*, and 'Knowledge of Other Minds', in *Knowledge and Certainty*, Englewood Cliffs, NJ: Prentice Hall. (Sets out Wittgenstein's alleged refutation of solipsism and scepticism about other minds.)

McDowell, J. (1982) 'Criteria, Defeasibility, and Knowledge', *Proceedings of the British Academy*, 68: 455-79. (Provides good criticism of Malcolm's view.)

Wittgenstein, L. (1978) *Blue and Brown Books*, Oxford: Blackwell.(Pages 23-5 are especially important.)

Wittgenstein, L. (1992) *Last Writings on the Philosophy of Psychology: The Inner and the Outer*, Oxford: Blackwell.(Provides the most sustained treatment of the topic of criteria. But note that it is in the nature of Wittgenstein's philosophical method that remarks on criteria are spread throughout his works and that these remarks bear complex relations both with each other and with other topics discussed.)

Wright, C. (1982) 'Anti-Realist Semantics: The Role of *Criteria*', in G. Vesey (ed.) *Idealism Past and Present*,

Cambridge: Cambridge University Press.(Provides an alternative to Malcolm's view.)

Deconstruction

Although the term is often used interchangeably (and loosely) alongside others like 'post-structuralism' and 'postmodernism', deconstruction differs from these other movements. Unlike post-structuralism, its sources lie squarely within the tradition of Western philosophical debate about truth, knowledge, logic, language and representation. Where post-structuralism follows the linguist Saussure - or its own version of Saussure - in espousing a radically conventionalist (hence sceptical and relativist) approach to these issues, deconstruction pursues a more complex and critical path, examining the texts of philosophy with an eye to their various blindspots and contradictions. Where postmodernism blithely declares an end to the typecast 'Enlightenment' or 'modernist' project of truth-seeking rational enquiry, deconstruction preserves the critical spirit of Enlightenment thought while questioning its more dogmatic or complacent habits of belief. It does so primarily through the close reading of philosophical and other texts and by drawing attention to the moments of 'aporia' (unresolved tension or conflict) that tend to be ignored by mainstream exegetes. Yet this is not to say (as its detractors often do) that deconstruction is a kind of all-licensing textualist 'freeplay' which abandons every last standard of interpretive fidelity, rigour or truth. At any rate it is a charge that finds no warrant in the writings of those - Jacques Derrida and Paul de Man chief among them - whose work is discussed below.

1 Is there a logic of deconstruction?

What is 'deconstruction' and what does it do? There is a problem here in that Derrida often goes out of his way to disown any summary treatment of the topic or any attempt to define (and so delimit) deconstruction in 'adequate' conceptual terms. Nevertheless one can offer at least some attempt at a working definition. What typically occurs in a deconstructive reading is that the text in question is shown to harbour contradictory logics which are standardly ignored - or concealed from view - on other, more orthodox accounts. Very often it is a matter of locating certain clearly-marked binary oppositions (as for instance between nature and culture, speech and writing, concept and metaphor, or philosophy and literature) and showing that their order of priority is by no means as stable or secure as the text seeks to maintain. That is to say, there is a counter-logic at work whereby those distinctions can be shown to break down, or to generate a reading markedly at odds with the author's overt intent. Not that intention is simply ruled out as irrelevant for the purposes of a deconstructive reading. On the contrary, it offers an 'indispensable guardrail' (Derrida [1967b] 1975: 158) which saves interpretation from running wild in endless subtleties of its own ingenious devising. However, this leaves open the possibility that texts may mean something other - and more - than is allowed for by any straightforward appeal to the warrant of authorial intention.

Thus in each of the above-mentioned cases one term (the first) is conventionally assigned a positive or superior value while the other is construed - by author and mainstream interpreters alike - as self-evidently standing in a dependent, derivative or supplementary relation to it. For Rousseau, nature both precedes culture and represents a better, more authentic way of life, while writing is regarded as a bad 'supplement' to speech, one that involves all manner of corrupting artifice and which infects the very sources of spontaneous 'natural' expression. Yet he cannot describe that idealized 'state of nature' except in terms that perforce acknowledge the *impossibility* of any such state having ever existed. For there is clearly a logical or conceptual problem about an argument that sets up this rigid opposition between nature and culture, but which then proceeds to identify the best (that is, the most primitive, least 'civilized') cultures with an early stage of social evolution that would somehow *not yet* have embarked upon the path to civilization and ruin. The case with speech and writing is somewhat more complex but can be stated along much the same lines (see Rousseau, J. §2).

For Rousseau, speech - in its 'natural' condition - is a mode of utterance that expresses sentiments so directly (with so little room for civilized pretence) that it allows human beings to communicate face-to-face without any need for the kinds of elaborate convention that characterize other, more articulate forms of linguistic exchange. What is bad about writing is precisely its dependence on a graphic notation which involves such artifice as the very condition of its possibility, and which thus stands at the furthest remove from the passional origins of speech. However this argument fails to take account of the conventional character of *all* language, spoken and written alike; that is to say, the extent to which speech partakes of various signifying systems (phonetic, semantic, syntactic and so on) in the absence of which it would simply not qualify as language. Furthermore it is just this reliance on conventions -

a matter of structural necessity - that Rousseau has to recognize (against his own wishes) as an intrinsic condition of language-in-general. For it would then be the case that writing (not speech) might offer the best model for explaining those constitutive features of language - its structural properties, conventional character, capacity to signify in various contexts irrespective of self-present utterer's intentions, and so forth - which necessarily pertain to *all* language, whether written or spoken (Derrida 1989).

There is a parallel here with Wittgenstein's argument against the possibility of a private language, that is to say, his contention that any such 'language' would not in fact constitute a language since it would lack the criteria for shared understanding and hence be wholly unintelligible even to its solitary user (see Private language argument; Wittgenstein, L.). What is distinctive about Derrida's approach to such issues is also what sets deconstruction apart from most work in the 'other' (analytic or Anglo-American) tradition. That is, he treats them as issues specific to the reading of certain problematical passages in certain texts, passages whose overt or manifest sense is shown to be at odds with the structural logic that governs the argument as a whole. Hence Derrida's claim with regard to Rousseau that 'his declared intention is not annulled by this but rather *inscribed* within a system which it no longer dominates' (Derrida [1967b] 1975: 243). However this emphasis on textual close-reading goes along with a keen analytical grasp of the philosophic issues involved and should not, therefore, be taken as evidence that Derrida is practising literary criticism (or a mode of 'merely' rhetorical exegesis) under the guise of linguistic philosophy. Rather he argues from the *conditions of possibility* for raising such questions about meaning, context, authorial intention and so forth, questions that clearly point beyond the sphere of localized interpretative insight or thematic commentary. Thus he is not so much concerned with particular instances of paradox, aporia or self-contradiction in Rousseau's texts. What chiefly engages Derrida's interest is that pervasive 'logic of supplementarity' whereby the second (supposedly inferior or derivative) term in each pair turns out to be always already presupposed in any definition of the first (supposedly original and self-sufficient) term. This logic may resist formalized statement through its complexities of modal structure - which Derrida brings out to striking effect in his reading of Rousseau - and its apparent suspension of classical axioms like bivalence or non-contradiction. But this should hardly place his work beyond the pale for philosophers with a knowledge of developments in recent post-classical (for example, intuitionist or many-valued) logic.

2 Paul de Man and 'aesthetic ideology'

Paul De Man is another thinker whose writing occupies the currently contested zone between philosophy and literary theory. In de Man's case also, the suspicion is often voiced that his practice of minute close-reading - especially when applied to philosophic texts - must go along with a lack of genuine analytic rigour or merely a desire to make trouble by inventing all manner of pseudo-problems. I should not wish to claim that this characterization is entirely wide of the mark, at least as regards de Man's more tortuous attempts to outdo Nietzsche in deconstructing all the concepts and categories of logic, epistemology, causal explanation and other 'metaphysical' notions (de Man 1979). All too often these claims have been taken on trust or treated as a deconstructive *fait accompli* by literary theorists keen to upstage their colleagues in philosophy or the natural sciences. At its best, however, de Man's work cannot be dismissed as simply a product of the turf-battle between rival academic disciplines. In late essays like 'The Epistemology of Metaphor' (1978) on Locke, Condillac and Kant, and 'Pascal's Allegory of Persuasion' (1981) on the relation between suasive or performative language and the language of logical proof-procedures, his thought exhibits an exceptional degree of analytic acuity and rigour. Certainly he is far from advancing the sorts of blanket 'textualist' claim - such as 'all concepts are metaphors', 'all philosophy is a kind of literature', 'all talk of "truth" is just a species of rhetorical imposture', and so forth - that are often attributed to him (and to Derrida) by disciples and opponents alike.

Above all, de Man insists that textual exegesis in the deconstructive mode can at times produce readings sharply at odds with established or mainstream critical opinion. This produces an extreme - some would say perverse - attentiveness to just those kinds of aberrant textual detail (tropological swerves, displacements, substitutions or elisions) which complicate the reading-process beyond any notion of straightforward fidelity to the author's presumed intent. Of course this description might also apply to that mode of 'rational reconstruction' by which some philosophers have lately seen fit to treat the texts of past thinkers with a view to their own, mainly analytic concerns. However it is the aim of this approach to discover some viable (constructive) solution to problems whose importance is taken to justify the strong-revisionist line. For de Man, conversely, it is the purpose of a deconstructive reading to respect the very letter of the text even where this leads to a stage of aporetic deadlock

where the text turns out to contest or subvert its own statements of intent.

Hence his insistence on the stubborn *literality* of deconstruction, that is to say, its power to hold out against the seductive ('eudaimonic') appeal of other, more compliant and therefore - he would argue - less rigorous modes of understanding. For de Man much depends on the way that we read certain passages in the work of philosophers and literary theorists who have reflected - in a more or less self-conscious or critical mode - on these issues of language, knowledge and representation. What he often discovers is an alternating rhythm of co-implicated 'blindness' and 'insight', a rhythm characterized by moments of apparent naïvety (for example, with regard to the supposed organic or consubstantial relation between language and nature) which none the less exist in close proximity to moments of un-self-deceiving rigour (de Man 1983). His aim is to expose the workings of a deep-laid 'aesthetic ideology' whose effect - so he argues - is chiefly manifest in the discourse of post-Kantian philosophy and literary theory. Most often it results from a confusion between linguistic (especially metaphorical) structures and those forms of phenomenal or sensory perception that characterize our knowledge of natural-world objects and processes. Thus : 'literature is fiction not because it somehow refuses to acknowledge "reality", but because it is not *a priori* certain that language functions according to principles which are those, or which are *like* those, of the phenomenal world' (de Man 1986: 11). To suppose otherwise is to fall into a way of thinking that can all too easily be co-opted for the purposes of ideological mystification.

Such is, for instance, the Cratylist delusion - revived by Romantic poet-philosophers like Schiller and Coleridge - according to which it is possible for language (especially the language of symbol and metaphor) to achieve a condition of organic form beyond all the Kantian antinomies of subject and object, mind and nature, or intuitive and conceptual knowledge. Aesthetic ideology can thus be seen as the source of some far-reaching errors, among them - most crucially for de Man - the idea that particular (national) languages and cultures can likewise manifest the kind of predestined, organic development and growth that grants them an inherent superiority over other, more fractured or dissociated forms of cultural-linguistic expression. De Man's notorious wartime journalism, published in Nazi-occupied Belgium, often attributed just such a natural, predestined eminence to German thought and culture. It seems plausible to claim that his later writings were concerned in various ways to challenge or resist what he subsequently saw as the seductive power of an aesthetic creed with large implications for our thinking about issues of culture, politics and moral responsibility. Thus: 'what we call ideology is precisely the confusion of linguistic with natural reality, of reference with phenomenalism'. And again, more pointedly, those who charge deconstruction with a culpable indifference to issues of history and politics 'are merely stating their fear at having their own ideological mystifications exposed by the tool they are trying to discredit. They are, in short, very poor readers of Marx's *German Ideology*' (de Man 1986: 11).

3 Some analytic bearings

Gilbert Ryle lost patience with the so-called 'psychologism' of Husserl, despite having previously published some detailed and constructive (albeit critical) studies of his work (Ryle 1971 - see Ryle, G. §1) . So it was - or so the story is often told - that a rift emerged between phenomenology (along with various derivative 'continental' schools of thought) and later developments in the Anglo-American analytic tradition. This view fitted in with the received idea - initiated by Russell and Frege, and erected into a full-scale programme by logical empiricists like Rudolf Carnap - that philosophy should work to clear away those errors and confusions of natural language that could best be revealed by analysis in the formal or logico-semantic mode (see Analytical philosophy §2). It would also help to resist the kinds of metaphysical bewitchment by language to be seen at their worst (so Carnap thought) in Heidegger's obscurantist appeal to the primordial wisdom supposedly enshrined in Greek and German etymology. Derrida's work has indeed been much influenced by Heidegger, especially as regards its questioning of those various deep-laid 'logocentric' priorities and values whose presence in the Western philosophical tradition is a constant theme of his writing, early and late. However he has also criticicized some aspects of Heidegger's thought, among them precisely the harking-back to a lost originary plenitude of Being (Derrida 1967b). For it is here that Heidegger himself falls prey to a potent irrationalist mystique, an appeal to language - or to certain privileged *national* languages and cultures - as the source of a wisdom supposedly concealed throughout the history of so-called 'Western metaphysics' (see Heidegger, M. §6).

In de Man's work also, Heidegger figures as a prime example of that alternating pattern of 'blindness' and 'insight' which in turn can enable the reader to attain a less prejudiced (if never fully adequate, transparent or

demystified) level of understanding (de Man 1983). Such arguments will scarcely be convincing to those for whom 'deconstruction' is pretty much synonymous with the obscurantist 'jargon of authenticity' that Heidegger promoted and that Derrida has supplied with its latest fashionable twist. However this notion cannot stand up to the sort of close reading which Derrida brings to his commentary on Heidegger and others, and which his own texts properly require on the part of responsible critics. Here again what has most often got in its way is the fixed preconception - handed down from the logical empiricists - that analytic (that is, genuine or serious) thought should have no truck with such extravagant forms of 'continental' pseudo-philosophy.

If the prospects look fair for a mending of this rift, then one clear sign is the current reappraisal (by Michael Dummett (1993) among others) of the strong continuities that exist within and between these two lines of descent. Another is the willingness of some commentators to look again, for example, at Derrida's reading of J.L. Austin - the target of a well-known hostile response by John Searle (1977) - and to discover not only points of genuine philosophical interest but also a certain affinity between the thought of Austin and Derrida (Austin 1961, Derrida 1989). This emerges, for instance, in their shared alertness to the ways that language (so-called 'ordinary' language) can prove more complex and resistant to codification than anything allowed for by speech-act theory in its systematized form. If 'undecidability' is in question here - as so often in Derrida's work - then it is the upshot of a reading closely attentive to the detail and the logic of Austin's arguments.

See also: Post-structuralism

CHRISTOPHER NORRIS

References and further reading

Austin, J.L. (1961) *How to Do Things With Words*, London: Oxford University Press.(A classic exposition of speech-act theory, 'deconstructed' with relish - though by no means attacked or dismissed - by Derrida in *Limited Inc.*)

de Man, P. (1978) 'The Epistemology of Metaphor', *Critical Inquiry* 5 (1): 13-30. (On metaphor in the texts of Locke, Condillac and Kant.)

de Man, P. (1979) *Allegories of Reading: figural language in Rousseau, Nietzsche, Rilke, and Proust*, New Haven, CT: Yale University Press.(Essays in his typical late mode of deconstructive textual exegesis.)

de Man, P. (1981) 'Pascal's Allegory of Persuasion', in S.J. Greenblatt (ed.) *Allegory and Representation*, Baltimore, MD: Johns Hopkins University Press, 1-25.(Analyses the relationship between constative and performative language in Pascal's writings on geometry, mathematics and religious faith.)

de Man, P. (1983) *Blindness and Insight: essays in the rhetoric of contemporary criticism*, London: Methuen, 2nd edn.(Examines the blind-spots of presupposition and ideological prejudice in various twentieth-century critics and philosophers.)

de Man, P. (1986) *The Resistance to Theory*, Minneapolis, MN: University of Minnesota Press.(On 'aesthetic ideology' and the resistance to it through a deconstructive reading of various - mainly literary-critical - texts.)

Derrida, J. (1967a) *La Voix et le Phénomène*; trans. D. Allison as *Speech and Phenomena and Other Essays on Husserl's Theory of Signs*, Evanston, IL: Northwestern University Press, 1973. (Represents Derrida's early, intensive, and critical engagement with Husserl.)

Derrida, J. (1967b) *De la Grammatologie*; trans. G. Spivak as *Of Grammatology*, Baltimore, MD: Johns Hopkins University Press, 1975.(On Rousseau, speech/writing, nature/culture, and other key themes of Derrida's early work.)

Derrida, J. (1967c) *L'Écriture et la Différence*; trans. A. Bass as *Writing and Difference*, Chicago, IL: The University of Chicago Press, 1978.(Essays on Husserl, Levinas, Foucault, structuralism and the human sciences.)

Derrida, J. (1972a) *Marges de la Philosophie*; trans. A. Bass as *Margins of Philosophy*, Chicago, IL: University of Chicago Press, 1982.(Contains some of Derrida's most important and powerfully-argued essays, among them 'White Mythology: metaphor in the text of philosophy'.)

Derrida, J. (1972b) *La Dissémination*; trans. B. Johnson as *Dissemination*, London: Athlone Press, 1981. ('Plato's Pharmacy' and other essays in the border-zone between philosophy and literature.)

Derrida, J. (1989) *Limited Inc*, ed. G. Graff, Evanston, IL: Northwestern University Press. (Includes Derrida's essay on Austin; also his response to Searle (1977) and further reflections on speech-act philosophy.)

Dummett, M. (1993) *On the Origins of Analytic Philosophy*, Cambridge, MA: Harvard University Press. (Comparative/contrastive account of work in the two traditions, with reference mainly to Frege and Husserl.)

Norris, C. (1987) *Derrida*, London: Fontana.(Introductory work for students of philosophy and literary theory.)

Norris, C. (1988) *Paul de Man and the Critique of Aesthetic Ideology*, London: Routledge. (Critical survey of de Man's writings; develops some of the arguments outlined above.)

Ryle, G. (1971) *Collected Papers*, vol. 1, London: Hutchinson.(Includes Ryle's early review-articles on Husserl and Heidegger.)

Searle, J.R. (1977) 'Reiterating the Differences', *Glyph*, vol. 1, Baltimore, MD: Johns Hopkins University Press. (Highly critical response to Derrida on the topic of Austinian speech-act philosophy.)

Demonstratives and indexicals

Demonstratives and indexicals are words and phrases whose interpretations are dependent on features of the context in which they are used. For example, the reference of 'I' depends on conditions associated with its use: as you use it, it refers to you; as I use it, it refers to me. In contrast, what 'the inventor of bifocals' refers to does not depend on when or where or by whom it is used. Among indexicals are the words 'here', 'now', 'today', demonstrative pronouns such as 'this', reflexive, possessive and personal pronouns; and compound phrases employing indexicals, such as 'my mother'. C.S. Peirce introduced the term 'indexical' to suggest the idea of pointing (as in 'index finger').

The phenomenon of indexicality figures prominently in recent debates in philosophy. This is because indexicals allow us to express beliefs about our subjective 'place' in the world, beliefs which are the immediate antecedents of action; and some argue that such beliefs are irreducibly indexical. For example, my belief that I am about to be attacked by a bear is distinct from my belief that HD is about to be attacked by a bear, since my having the former belief explains why I act as I do (I flee), whereas my having the latter belief explains nothing unless the explanation continues 'and I believe that I am HD'. It seems impossible to describe the beliefs that prompt my action without the help of 'I'. Similarly, some have argued that indexical-free accounts of the self or of consciousness are necessarily incomplete, so that a purely objective physicalism is impossible. In a different vein, some (such as Putnam 1975) have argued that our terms for natural substances, kinds and phenomena ('gold', 'water', 'light') are indexical in a way that entails that certain substantive scientific claims - for example, that water is H_2O - are, if true, necessarily true. Thus, reflection on indexicality has yielded some surprising (and controversial) philosophical conclusions.

1 Semantic data

An indexical is context-dependent: its interpretation varies with changes in aspects of the context of use. Yet almost any expression exhibits variation in interpretation with respect to some parameter or other. For example, a contingent sentence is one which is possibly true and possibly false, so we may think of it as varying in truth-value with changes in possible worlds. But mere contingency is not a species of indexicality. Similarly, tensed verbs induce variation in interpretation with respect to the passage of time, but tense is not usually viewed by linguists and logicians as a kind of indexicality (although there is some confusion about the matter). The difference between true context dependence and mere 'parameter dependence' can be seen as follows. Consider the sentence

(1) Jones will remember what Smith said just now.

The truth or falsity of this sentence at a moment in time t depends on the truth or falsity of the present tense sentence

(2) Jones remembers what Smith said just now

at moments t' future to t. But the referent of the occurrence of 'now' in (2) must refer back to t and not to any of the future moments t'. This means that the semantic rule for 'now' cannot be simply the rule which states that at any time t, 'now' refers to t. Instead, the rule must say that a use of 'now' at a time t will refer at time t *and at any other time* to t. Thus, the rule for 'now' must be 'double-indexed'. In general, the semantic rules for indexicals must involve the pairing of two sorts of parameter: contextual and circumstantial. The contextual parameters (speaker, time of utterance, place of utterance, world of utterance, addressee and so on) fix the referent of an indexical, so that its interpretation is unaffected by any circumstantial parameter (possible world, moment of time) required by the presence of tense or modal elements. Hereafter, the phrase 'context-dependent' will refer to the dependence of indexicals on both context and circumstance.

Indexicals are 'referential' in that they invariably have wide scope when embedded in tensed, modal or propositional attitude constructions. For example, consider the two statements 'The mayor might not have been a Democrat' and 'That person over there, the Mayor, might not have been a Democrat'. The former statement could be interpreted to mean merely that the office of mayor might have been occupied by a non-Democrat; but the latter statement can only be interpreted to mean that a certain person (who happens to be the mayor) might not have been

a Democrat. This property of indexicals is an effect of their context dependence. The referent of the indexical phrase 'that person' is fixed by the context and not by circumstances required to interpret the modal element 'might have been'.

Though the interpretation of an indexical will vary with the context, indexicals are not in general ambiguous. Your use of 'I' and mine are linked by a common meaning. The language supplies a rule - the same rule for both of us - for the use of 'I'. The rule is roughly this: 'As *x* uses "I" in such and such a context, "I" refers to *x* in all circumstances'. But this sort of meaning is not like the Fregean sense of a definite description or proper name; it is not a concept of some fixed object and it cannot be combined with the sense of a (non-indexical) predicate to yield a complete proposition (see Frege, G. §§2-3). Some theorists believe that this fact requires the introduction into semantics of a whole new level of meaning distinct from both sense and reference. Kaplan, for example, draws a three-way distinction between the 'character' (linguistic meaning), content (roughly, sense) and referent of an indexical expression such as 'my mother' (1970). Kaplan's view may not be inevitable, however. It may be that 'I' has an 'incomplete sense', in the way that the sense of a functional expression 'the square root of *x*' or 'the author of *x*' is incomplete. If so, that would explain why the linguistic meaning of 'I' cannot be combined with a predicate to yield a complete proposition and there would be no need to postulate a kind of meaning in addition to sense and reference.

Some subscribe to the principle that for one to understand an 'object-dependent' proposition - a proposition that such and such an object has such and such a property - one must be able to identify the object in question; one must be able to say who or what it is. This principle seems to be directly contradicted by propositions involving indexicals. I can know that I am hungry or that it is raining now without having any substantive information about who I am or what time it is. Similarly, if I point to something and say 'I don't have the slightest idea what that is' I might well speak the truth! Yet these propositions are object-dependent in that they would not be available if the objects they are about did not exist. Let us mark the fact that speakers can use indexicals with total understanding without having any substantive knowledge of their referents by saying that indexicals are 'transparent'.

Some indexical sentences have the property of being true in any context of use. For example, one who says (or thinks) 'I exist' or 'I am here now' or 'I am saying (or thinking) something now' must be right. These sentences, then, are somewhat similar to logical truths. But if I say 'I exist', then what I say (that HD exists) is not a *necessary* truth, for there are possible circumstances in which I do not exist. So a sentence such as 'I exist' is contingent in one respect and yet non-contingent in another. Any theory of indexicality must explain how this is possible.

Perhaps the most puzzling aspect of indexicals is that some of them seem to be ineliminable and their ineliminability seems to have to do with the baffling distinction between subjectivity and objectivity (see Content, indexical). Perry gives the following example (1979). Suppose a meeting is set for Tuesday at noon. All week you believe that the meeting will start at noon on Tuesday. At a certain point, you leave to go to the meeting. Obviously, you must have acquired some new information, a new belief, distinct from the belief that the meeting starts at noon on Tuesday, for you now act in a way you have not acted all week (that is, you leave to go to the meeting). What belief is this? It is the belief that the meeting starts *now*, or, more elaborately, the belief that the meeting starts at noon on Tuesday and it is *now* noon on Tuesday. The use of 'now' in describing your new belief seems essential, and not eliminable in favour of non-indexical designations of noon on Tuesday. Moreover, your new 'indexical belief' places you within the fabric of time in a way that the beliefs you acquire when you look at your appointment book to determine the time of the meeting do not. An observer 'outside of time' could have the belief that the meeting starts at noon on Tuesday, but not the belief that the meeting starts 'now'. In this way, your indexical belief seems linked to your subjective place in the world. Perry calls such beliefs 'self-locating' beliefs. Yet all this is quite puzzling. Is your belief that it is 'now' noon on Tuesday something that you can believe only 'now'? Certainly, you can believe truly that it is now noon on Tuesday only at noon on Tuesday. But can you not at a later time continue to hold precisely the belief you expressed earlier by saying 'It is now noon on Tuesday'? Yet if you believe at the later time exactly what you believed at the earlier, then why do you not, at the later time, do just what you did at the earlier time - namely, head off to the meeting? Similarly, I believe that I am HD. Can you not believe that too? If you believe just what I do when I believe that I am HD - that is, if you believe exactly the proposition I do, it would appear to follow that our mental states *vis-à-vis* this belief must be identical, then why do you not have any inclination, as I do, to identify yourself as HD? Let us call this 'the puzzle about

indexical belief'. The puzzle has two parts: first, one argues on the basis of behavioural changes (for example, your getting up to go to the meeting) that an indexical belief (assertion, thought) is not equivalent to any non-indexical counterpart of it. Second, however, similar behavioural changes - or the lack of any - seem to show that such beliefs are 'isolated' - that they cannot be retained over time or shared by different persons: only I can know that I am HD, and that seems counterintuitive. Any adequate theory of indexicality must solve this puzzle about indexical belief.

2 The Kaplan-Perry theory

The most influential theory that has had some success explaining the semantic data about indexicals as outlined above was developed by Kaplan in a series of papers originating in the late 1960s and early 1970s. In the mid-1970s, Perry applied Kaplan's theory to explain the data about the ineliminability of indexicality and the puzzle about indexical belief - data that were first noticed and systematically studied by Castañeda (1967).

According to Kaplan-Perry theory (KPT), each indexical expression has a 'character' or linguistic meaning (Perry's term is 'role') and, given a context, a 'content' in that context. The content in a context of a (declarative) sentence containing an indexical is a 'singular proposition' - a proposition containing the very referent of the indexical as a 'constituent' rather than a sense or concept of that referent. So the content of my use of 'I am hungry now' at a time t is a proposition to the effect that HD is hungry at time t - with HD right there 'in' the proposition, in at least a metaphorical sense. Kaplan takes the character of an indexical to be a function (in the mathematical sense) that takes context into content. Contexts are treated as n-tuples $c = \langle a, t, p, w, \ldots \rangle$ with coordinates for the 'agent of the context' (to handle 'I'), the 'time of the context' (to handle 'now'), the 'place of the context', the 'world of the context' and so on (for example, there should be a coordinate for the 'day of the context' to handle 'yesterday', 'today' and 'tomorrow'). To deal with pure demonstratives such as 'this', Kaplan adds a coordinate for a 'demonstration'. This is a kind of abstract counterpart of an act of pointing. For example, the sentence 'That is mine now' will be true at a parameter $\langle a, t, p, w, d, \ldots \rangle$ if the demonstrated object d belongs to agent a at t in place p and world w.

Kaplan distinguishes between 'contexts' and 'circumstances'. Formally, the latter are pairs $\langle t, w \rangle$ with t a time and w a possible world. Strictly, expressions are evaluated relative to parameters of the form $\langle \mathrm{context, circumstance} \rangle$. So the semantic rules are double-indexed. In addition, Kaplan claims that indexicals are 'directly referential'. By this he means that an indexical refers without the aid of a Fregean sense. In this respect indexicals are like proper names. But proper names are, in Kaplan's view, unlike indexicals in that the latter possess characters whereas the former do not (see Proper names §§2-3).

By appealing to the content/character and context/circumstance distinctions, KPT can readily explain the fact that indexicals are context-dependent, referential and univocal. Moreover, the character/content distinction helps to explain the transparency of indexicals. One need only grasp the character of an indexical to use it with understanding. On occasion, its content may be hidden, as in the case of amnesiacs, who, despite their affliction, can certainly use the first person pronoun perfectly competently. KPT can also explain the special status of sentences such as 'I exist'. The character of 'I exist' always produces a true content in any context. But that content may turn out to be false in further circumstances. For example, if you now write 'I exist' in your diary, the content you express is true now. But, inevitably, that content will be false years hence in circumstances in which you no longer exist and someone then reads these words in your diary. According to KPT, 'I exist' is a logical truth, although its necessitation, 'Necessarily, I exist', and its temporal necessitation, 'It is always the case that I exist', are not. KPT thus involves a novel conception of logical truth. For it is usually assumed that the class of logical truths contains 'Necessarily, *P*' and 'It is always the case that *P*' if it contains *P*. Finally, KPT solves the problem of indexical belief by appealing to the notion of character. KPT holds that belief is linked to action via character, not content. If you and I have beliefs with the same content but different character we may well behave differently. But if we have beliefs with the same character and different content we may well behave in the same way. When each of us thinks 'I am about to be attacked by a bear' (same character, different content) we both run for our lives, but when I think 'I am about to be attacked by a bear' and you think 'You are about to be attacked by a bear' (same content, different character) I run for my life and you do not. Similarly, you can retain the belief that the meeting starts now but you can express it at a later time only by using words having a character distinct from that of 'The meeting starts now', and it is the difference in character that accounts for the fact that, at the later

time, you do not do what you did at the earlier time. Likewise, KPT appeals to character to account for the difference between the thought that I am about to be attacked by a bear and the thought that HD is about to be attacked by a bear. The former has a certain character which the latter does not, and that makes all the difference. Similarly, consider the difference between the thought that I am HD and that HD is HD. The former is informative whereas the latter is not. Again, the difference can be traced to the fact that 'I' has a certain character which 'HD' does not.

KPT needs only the character/content and context/circumstance distinctions to account for the semantic data. But it goes further in claiming that indexicals are 'directly referential' and that indexicals give rise to 'singular propositions'. Proponents of KPT argue in several ways for the direct reference thesis. First, Kaplan argues that 'I' does not have the sense of 'the speaker' or any description such as 'the utterer' or 'the thinker'. He would argue, for example, that since 'I am not the speaker' is not contradictory, whereas 'The speaker is not the speaker' is contradictory, 'I' and 'the speaker' are not synonymous. But this argument seems equivocal. Any candidate synonym for 'I' must be context-dependent (and not just parameter-dependent). If we associate a context-dependent rule with 'the speaker' - so that it refers in any circumstance to the 'speaker of the context' - then 'I am not the speaker' is indeed contradictory. It is only by treating 'I' as context-dependent and 'the speaker' as merely parameter-dependent (as it is in one sense) that there is a discrepancy. Second, Kaplan argues in the following way. The descriptive meaning of 'now' is given by the following rule: at t, 'now' refers to t. Thus, the descriptive meaning of 'now' is 'either inapplicable or irrelevant to determining a referent with respect to a circumstance of evaluation' (1989: 500). Hence, 'now' is directly referential - its descriptive meaning does not serve as the route to its referent in a circumstance. This argument does not work either. One can take the view, as explained above, that the descriptive rule for 'now' is not the one Kaplan assumes. (So the first premise of Kaplan's argument is false.) And if we substitute the double-indexed rule for 'now' it cannot be argued that descriptive meaning (that is, the double-indexed rule) of 'now' is 'irrelevant or inapplicable' to determining its referent in a circumstance.

See also: Logical and mathematical terms, glossary of;Reference §5

HARRY DEUTSCH

References and further reading

Braun, D. (1995) 'What is Character?', *Journal of Philosophical Logic* 24: 227-40.(A critical and relatively nontechnical discussion of some aspects of Kaplan's notion of character.)

Castañeda, H.-N. (1967) 'Indicators and Quasi-Indicators', *American Philosophical Quarterly* 4: 85-100.(This is a representative example of several pioneering papers by this author exploring the behaviour of indexicals in propositional attitude constructions.)

Forbes, G. (1989) *Languages of Possibility*, Oxford: Oxford University Press.(This work includes a detailed neo-Fregean theory of indexicality. In contrast, the Kaplan-Perry theory discussed above may be described as 'anti-Fregean'.)

Kamp, H. (1971) 'Formal Properties of "Now"', *Theoria* 37: 227-73.(This paper develops a double-indexed semantics for certain tense logics with an operator representing 'now'. Kamp was the first to recognize the need for the double-index technique.)

Kaplan, D. (1970) 'On the Logic of Demonstratives', *Journal of Philosophical Logic* 8: 81-98.(The pioneering paper in which the key distinctions between character and content and context and circumstance were first introduced.)

Kaplan, D. (1989) 'Demonstratives', in *Themes from Kaplan*, ed. J. Almog, J. Perry and H.K. Wettstein, Oxford: Oxford University Press.(This is a classic meditation on the anti-Fregean aspects of the Kaplan-Perry theory.)

Perry, J. (1977) 'Frege on Demonstratives', repr. in P. Yourgrau (ed.) *Demonstratives*, Oxford: Oxford University Press, 1990.(This paper attempts to demonstrate the inability of Frege's semantics to handle indexicality and the inadequacy of Frege's own attempts to deal with the phenomenon.)

Perry, J. (1979) 'The Problem of the Essential Indexical', *Noûs* 13: 3-21.(By means of several ingenious examples this paper uncovers and attempts to solve the problem of indexical belief.)

Putnam, H. (1975) 'The Meaning of "Meaning"', in *Mind, Language and Reality: Philosophical Papers*, vol. 2, Cambridge: Cambridge University Press.(This paper is the main source of the view according to which our terms for natural substances, kinds and phenomena are indexical.)

Yourgrau, P. (ed.) (1990) *Demonstratives*, Oxford: Oxford University Press.(This is a collection of essays on

indexicality which emphasize not only semantic issues but also metaphysical and epistemological themes.)

Descriptions

'Definite descriptions' are noun phrases of the form 'the' + noun complex (for example, 'the finest Greek poet', 'the cube of five') or of the form possessive + noun complex (for example, 'Sparta's defeat of Athens'). As Russell realized, it is important to philosophy to be clear about the semantics of such expressions. In the sentence 'Aeschylus fought at Marathon', the function of the subject, 'Aeschylus', is to refer to something; it is a referential noun phrase (or 'singular term'). By contrast, in the sentence 'Every Athenian remembers Marathon', the subject noun phrase, 'every Athenian', is not referential but quantificational. Definite descriptions appear at first sight to be referential. Frege treated them referentially, but Russell held that they should be treated quantificationally in accordance with his theory of descriptions, and argued that certain philosophical puzzles were thereby solved.

1 Frege

Gottlob Frege provided the first systematic account of quantification in natural language and the first systematic theory of reference (1892). The class of 'singular terms' (referential noun phrases), for Frege, was delimited by a set of logical tests (for example, the licensing of existential generalization) and was recursive. It included ordinary proper names and definite descriptions. Thus '5', 'the cube of 5', 'Aeschylus' and so on were all singular terms.

If a description ⌜the F⌝ is referential, then it is natural to take its reference to be the unique entity satisfying 'F'; a sentence ⌜The F is G⌝ is true if and only if that entity is G. But what if no entity, or more than one entity, satisfies the descriptive condition, as in (1) or (2)?

(1) *The largest prime number* lies between 10^{23} and 10^{27}.
(2) *The man who landed on the moon* was American.

Such descriptions are said to be 'improper'. Frege considered it a defect of natural language that it permits the possibility of improper terms. As far as his own logical system was concerned, he thought it essential that every formula have a truth-value, and so he insisted that every singular term have a reference (or meaning): he stipulated that a specified object in the range of the quantifier(s) serve as the referent of every improper description. While this stipulation proved useful for his formal system, Frege recognized that some other account of improper descriptions in ordinary language was needed. Once he had made his distinction between the 'sense' and 'reference' of an expression, he suggested that an improper term has a sense but no reference (see Frege, G. §3; Sense and reference §1). The main problem with this proposal is that it predicts, rather counterintuitively, that any sentence containing an improper term (in a transparent context) lacks a truth-value. For example, it predicts that (1) and (2) above lack truth-values.

2 Russell's theory of descriptions: informal characterization

Like Frege, Bertrand Russell thought it important to explain how a sentence such as (1) or (2) could be meaningful. At one time he entertained the idea of a realm of non-existent entities to serve as the referents of descriptions such as 'the largest prime number' and 'the round square'; but by 1905 he thought this idea conflicted with a 'robust sense of reality' and his theory of descriptions came about, in part as an attempt to purify his ontology (see Existence §2).

On Russell's account, descriptions are not singular terms at all but phrases that logical analysis reveals to be quantificational: if ⌜the F⌝ is a definite description and ⌜ . . . is G⌝ is a predicate phrase, then the proposition expressed by an utterance of ⌜The F is G⌝ is equivalent, says Russell, to the proposition expressed by an utterance of ⌜There is exactly one F, and everything that is F is G⌝. That is, ⌜The F is G⌝ is analysed as

(3) $(\exists x)((\forall y)(Fy \leftrightarrow y = x) \,\&\, Gx)$.

The proposition expressed by ⌜The F is G⌝ is 'general' ('object-independent') rather than 'singular' ('object-dependent') in the sense that there is no object for which its grammatical subject stands; upon whose existence that of the proposition expressed depends. Unlike a singular term, a definite description, even if it is in fact satisfied by a unique object, does not actually *refer* to that object. It is as wrong, on Russell's account, to inquire into the referent of ⌜the F⌝ as it is to inquire into the referent of ⌜every F⌝ or ⌜no F⌝.

On Russell's account, sentences containing improper descriptions have truth-values. For example, (1) above is false as it is not the case that there exists a largest prime number. Similarly, sentence (2) is false as it is not the case that there exists exactly one man who landed on the moon.

Russell's theory opens up the possibility of accounting for certain *de dicto/de re* ambiguities in terms of scope permutations (see De re/de dicto). For example, (4) below may be analysed as either (5) or (6), according to whether the description 'the largest prime number' is given large or small scope with respect to 'John thinks that':

(4) John thinks that the largest prime number lies between 10^{23} and 10^{27}.
(5) $(\exists x)((\forall y)(\text{largest-prime } y \leftrightarrow y = x)$ & John thinks that: x lies between 10^{23} and $10^{27})$.
(6) John thinks that: $(\exists x)((\forall y)(\text{largest-prime } y \leftrightarrow y = x)$ & x lies between 10^{23} and $10^{27})$.

(5) is false; but (6) may be true. Thus Russell is able to explain the intuitive ambiguity of (4) and avoid positing an ontology that includes such things as a largest prime number. Smullyan (1948) points out that Russell's theory similarly explains *de dicto/de re* ambiguities in modal contexts, for example, in 'The number of planets is necessarily odd' (see De re/de dicto §2; Modal logic).

Russell came to treat ordinary proper names as 'disguised' or 'truncated' descriptions. For example, the name 'Cicero' might be analysed as the description 'the greatest Roman orator' while the coreferential name 'Tully' might be analysed as the description 'the author of *De Fato*'. On the face of it, this provided Russell with accounts (not dissimilar to Frege's) of why 'Cicero was bald' and 'Tully was bald' differ in informativeness, and of why (7) and (8) need not agree in truth-value:

(7) John believes Cicero was bald.
(8) John believes Tully was bald.

In the light of Kripke's work on names and necessity (1980), it is widely held that descriptive analyses of proper names cannot succeed (see Proper names). There is good reason, however, to think that at least some pronouns anaphorically linked to (referring back to), but not bound by, quantified noun phrases are understood in terms of definite descriptions (see Neale 1990).

3 Russell's theory of descriptions: formal characterization

On Russell's account, descriptions are 'incomplete' symbols; they have 'no meaning in isolation', that is, they do not stand for things. In *Principia Mathematica* (1910-13), descriptions are represented by quasi-singular terms of the form ' $(\iota x)(Fx)$ ', which can be read as 'the unique x which is F'. Superficially, the iota-operator is a variable-binding device for forming a term from a formula. A predicate symbol 'G' may be prefixed to a description ' $(\iota x)(Fx)$ ' to form a formula ' $G(\iota x)(Fx)$ ', which can be expanded in accordance with a suitable 'contextual definition'. (To define an expression ζ contextually is to provide a procedure for converting any sentence containing occurrences of ζ into an equivalent sentence that is ζ-free.)

The analysis in (3) above does not constitute a final contextual definition of ' $G(\iota x)(Fx)$' because of the possibility of scope ambiguity where a formula containing a description is itself a constituent of a larger formula (see Scope). Scope ambiguity is conveniently illustrated with descriptions in the context of negation. For a genuine singular term α, there is no difference between wide and narrow scope negation: α is not-F just in case it is not the case that α is F. For a description, however, there is a formal ambiguity. Let 'Kx' represent 'x is a king of France' and 'Wx' represent 'x is wise'. Then the formula ' $\neg W(\iota x)(Kx)$ ' ('The king of France is not wise') is ambiguous between (9) and (10):

(9) $(\exists x)((\forall y)(Ky \leftrightarrow y = x)$ & $\neg Wx)$
(10) $\neg(\exists x)((\forall y)(Ky \leftrightarrow y = x)$ & $Wx)$.

These are not equivalent: only (10) can be true when there is no king of France. In *Principia Mathematica*, the scope of a description is specified by appending a copy of it within square brackets to the front of the formula that constitutes its scope. Thus (9) and (10) are represented as (11) and (12) respectively:

(11) $[(\iota x)(Kx)]\neg\{W(\iota x)(Kx)\}$
(12) $\neg\{[(\iota x)(Kx)]W(\iota x)(Kx)\}$.

In (11) the description has what Russell calls a 'primary occurrence' by virtue of having scope over the negation; in (12) the description has a 'secondary occurrence' by virtue of lying within the scope of the negation. Where a description has smallest possible scope, it is conventional to omit the scope marker; thus (12) can be reduced to '$\neg W(\iota x)(Kx)$'.

With the matter of scope behind us, the theory of descriptions can be stated exactly:

$$[(\iota x)(\phi x)]G(\iota x)(\phi x) =_{\text{df}} (\exists x)((\forall y)(\phi y \leftrightarrow y = x) \,\&\, Gx),$$

where ϕ is a formula. On Russell's account, there is no possibility of a genuine referring expression failing to refer, so no predicate letter in the language of *Principia Mathematica* stands for 'exists'. Russell introduces a symbol 'E!' ('E shriek') that may be combined with a description to create a well-formed formula. Thus

$$E!(\iota x)(\phi x) =_{\text{df}} (\exists x)(\forall y)(\phi y \leftrightarrow y = x).$$

'E!' allows a treatment of negative existentials. (According to Russell, an utterance of 'The king of France does not exist' made today would be true precisely because there is no king of France.) Successive applications will allow any well-formed formula containing a definite description to be replaced by an equivalent formula that is description-free.

It is often objected that Russell's theory, which substitutes complex quantificational structure for 'the', is unfaithful to surface syntax. The objection is engendered by an insufficiently keen appreciation of the distinction between a theory and its formal implementation. The extent of the mismatch between 'The king is wise' and its analysis

(13) $(\exists x)((\forall y)(\text{king } y \leftrightarrow y = x) \,\&\, \text{wise } x)$

has nothing to do with descriptions *per se*. In order to characterize the logical forms of even 'some philosophers are wise' and 'every philosopher is wise' in the predicate calculus we have to use formulas containing sentence connectives, no counterparts of which occur in the surface forms of the sentences:

(14) $(\exists x)(\text{philosopher } x \,\&\, \text{wise } x)$
(15) $(\forall x)(\text{philosopher } x \rightarrow \text{wise } x)$.

And when we formalize sentences such as 'Just two philosophers are wise', we find much more complexity than there is in surface syntax:

(16) $(\exists x)(\exists y)[\text{philosopher } x \,\&\, \text{philosopher } y \,\&\, \text{wise } x \,\&\, \text{wise } y \,\&\, (\forall z)((\text{philosopher } z \,\&\, \text{wise } z) \rightarrow z = x \lor z = y)]$.

The supposed problem about descriptions, then, is in fact a symptom of a larger problem involving the application of first-order logic to sentences of ordinary language.

Work on 'generalized' quantification provides a solution to the larger problem (as well as treatments of quantifiers such as 'most' that cannot be handled within first-order logic; see Quantifiers, generalized). Natural language quantification is normally restricted: we talk about all philosophers or most poets, not about all or most entities. A simple modification of the predicate calculus yields a language - call it 'RQ' - that captures this fact while retaining the precision of regular first-order logic. In RQ, a determiner such as 'some', 'every' or 'no' combines with a formula to create a restricted quantifier such as '[every *x*: philosopher *x*]'. And such a quantifier may combine with a formula to form a formula:

(17) [every *x*: philosopher *x*] (wise *x*).

The viability of such a language shows that the language of *Principia Mathematica* is not essential to the theory of descriptions. Since the word 'the' is a one-place quantificational determiner (as are 'some', 'every', 'no' and so on), RQ can treat 'the' as combining with a formula 'king *x*' to form a restricted quantifier '[the *x*: king *x*]'. The sentence 'The king is wise' will then be represented as

(18) [the *x*: king *x*] (wise *x*).

Different scope possibilities are easily captured. For instance, 'The king is not wise' is ambiguous between (19) and (20):

(19) [the x: king x] $\neg$(wise x)
(20) $\neg$[the x: king x](wise x).

Using a formal language in which descriptions are treated as restricted quantifiers does not mean abandoning Russell's view that descriptions are 'incomplete symbols' that 'disappear on analysis'. Rather, treating descriptions as restricted quantifiers results in an explanation of where his theory of descriptions fits into a systematic account of natural language quantification, a theory in which 'every', 'some', 'most', 'a', 'the' and so on are members of a unified syntactic and semantic category.

4 Strawson's theory and criticisms of Russell

As part of a broad critique of the idea that the semantics of formal languages can be used to analyse the meanings of statements of natural language, P.F. Strawson argued against Russell's theory of descriptions on the grounds that (1) it fails to recognize that referring is something done by speakers and not expressions, (2) it fails to do justice to the way speakers ordinarily use sentences containing descriptions to make statements (speakers use descriptions to refer, not to quantify) and (3) it rides roughshod over important distinctions, such as the distinction between the meaning of a sentence σ and the statement made by a particular use of σ (see Propositions, sentences and statements).

Using as an example 'The present king of France is wise', Strawson argues that Russell's theory is thwarted because the same sentence can be used to say something true on one occasion and something false on another. It is certainly true that Russell paid little attention to the distinction between the linguistic meaning of a sentence type and the proposition expressed by a particular dated utterance of that sentence type; but it was the latter that actually concerned him, and Strawson could get no mileage out of Russell's inattention to the distinction. The fact that a description (or any other quantified noun phrase) may contain an indexical component ('the *present* king of France', 'every man *here*', and so on) illustrates that some descriptions are subject to both the theory of descriptions and a theory of indexicality (see Demonstratives and indexicals). Thus contextual features play a role in fixing the proposition expressed. And this can be true also if the overt indexical element is missing, as in 'The king of France is wise'.

This appreciation of contextual factors forms the basis of the Russellian response to a second Strawsonian objection. According to this, someone who uses a description ⌜the F⌝ typically intends to refer to some object or other and say something about it; there is no question of claiming that some object uniquely satisfies F. Someone who says 'The table is covered with books', for instance, does not express a proposition that entails the existence of exactly one table. But, Strawson claimed, it is a part of the meaning of ⌜the F⌝ that such an expression is used correctly only if there is an F. If this condition is not satisfied - if the 'presupposition' that there is an F is false - a use of ⌜The F is G⌝ cannot be considered to express a proposition that is either true or false. The Russellian response to Strawson is that descriptions such as 'the table' are often understood as elliptical uses of fuller descriptions such as 'the table over there', 'the table in front of me' and so on; or else they are evaluated with respect to a restricted domain of discourse. Again the phenomenon is not confined to descriptions, but is found with quantified noun phrases more generally.

Strawson's original statement (1950) of his own theory contains an interesting ambiguity. He can be understood as claiming either that no proposition is expressed, or that a proposition which is neither true nor false is expressed, when someone uses a sentence containing an empty description. A second ambiguity comes with the notion of 'presupposition'. This can be viewed as an epistemological or pragmatic relation between a person and a statement, or as a logical relation between two statements (see Presupposition). An epistemological or pragmatic notion of presupposition appears to have no bearing on the semantic issues Strawson wanted to address when he challenged Russell.

The Strawsonian position faces some serious obstacles. If someone were to utter (21) right now, they would unquestionably say something false.

(21) The king of France shot my cat last night.

But on Strawson's account, the speaker will have expressed no proposition because the presupposition that there is a unique king of France is false. Descriptions occurring in the context of attitude verbs create a similar problem. For example, someone might utter a true statement using (22):

(22) Ponce de León thought the fountain of youth was in Florida,

so the presence of an empty description does not always result in a failed speech act. This is something Strawson (1964) came to concede. In order to reduce the number of incorrect predictions made by his earlier theory, he suggested that the presence of an empty description sometimes renders the proposition expressed false and at other times prevents a proposition from being expressed at all (sometimes ⌜The F is G⌝ *entails* the existence of a unique *F*, and at other times it (only) *presupposes* it).

5 Ambiguity theories

Consideration of the behaviour of descriptions in non-extensional contexts and the possibility of misdescribing an individual as the *F*, but successfully communicating something about that individual, has led some philosophers (for example, Donnellan 1966) to suggest that descriptions are sometimes quantificational, at other times referential. When ⌜the F⌝ is used in the Russellian way, the proposition expressed is general (object-independent); when it is used referentially the proposition expressed is singular (object-dependent). Donnellan considers examples such as the following: (1) A detective discovers Smith's mutilated body but has no idea who killed him. Looking at the body, he exclaims, 'The murderer is insane'. (2) Jones is on trial for Smith's murder; I am convinced of his guilt; hearing Jones ranting in court, I say, 'The murderer is insane'. On Donnellan's account, in case (1) the description is being used attributively; in case (2) it is being used referentially. Cases such as (2), it is argued, cannot be treated in accordance with Russell's theory. Following Grice (1969), however, many have argued (1) that so-called referential uses of descriptions can usually be accommodated within Russell's theory by invoking a distinction between the proposition expressed by (an utterance of) a sentence on a given occasion and the proposition the speaker primarily intends to communicate on that occasion; (2) that the phenomenon of referential usage is not specific to definite descriptions, but arises with quantified noun phrases quite generally; (3) that the referential/attributive distinction is neither exclusive nor exhaustive; and (4) that no such distinction can do the work of Russell's notion of the scope of a description. It would seem, then, that something very close to Russell's theory will probably form a component of any finally acceptable theory.

See also: Free logics, philosophical issues in §3; Logical and mathematical terms, glossary of; Reference §6

STEPHEN NEALE

References and further reading

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press. (Discussion of views presented in §§1-2.)

Davies, M. (1981) *Meaning, Quantification and Necessity*, London: Routledge. (Discussion of views presented in §5.)

Donnellan, K. (1966) 'Reference and Definite Descriptions', *Philosophical Review* 77: 281-304.(Discussion of views presented in §5.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980.(Source of views presented in §1. The page numbers of the original publications appear in the text.)

Grice, H.P. (1969) 'Vacuous Names', in D. Davidson and J. Hintikka (eds) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 118-45. (Discussion of views presented in §§2, 5.)

Kaplan, D. (1978) 'Dthat', in P. Cole (ed.) *Syntax and Semantics*, vol. 9, *Pragmatics*, New York: Academic Press, 221-43.(Discussion of views presented in §5.)

Kripke, S.A. (1979) 'Speaker Reference and Semantic Reference', in P.A. French, T.E. Uehling and H.K. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, 6-27.(Discussion of views presented in §§2, 5.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(See §2.)

Neale, S. (1990) *Descriptions*, Cambridge, MA: MIT Press.(Discussion of views presented in §§2-5. Extensive bibliography.)

Neale, S. (1994) 'Grammatical Form, Logical Form, and Incomplete Symbols', in A.D. Irvine and G.A. Wedeking (eds) *Russell and Analytic Philosophy*, Toronto, Ont.: University of Toronto Press, 97-139.(Discussion of views presented in §2.)

Russell, B.A.W. (1905) 'On Denoting', *Mind* 14: 479-93; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992.(Source of views presented in §2.)

Russell, B.A.W. (1911) 'Knowledge by Acquaintance and Knowledge by Description', in *Mysticism and Logic*, London: Routledge, 1986.(Source of views presented in §2.)

Sainsbury, R.M. (1979) *Russell*, London: Routledge & Kegan Paul.(Discussion of views presented in §§2, 5.)

Sellars, W.S. (1954) 'Presupposing', *Philosophical Review* 63: 197-215.(Discussion of views presented in §4.)

Smullyan, A.F. (1948) 'Modality and Description', *Journal of Symbolic Logic* 13: 31-7; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971, 35-43.(Discussion of views presented in §2.)

Strawson, P.F. (1950) 'On Referring', *Mind* 59: 320-44.(Source of views presented in §4.)

Strawson, P.F. (1954) 'Reply to Mr. Sellars', *Philosophical Review* 63: 216-31.(Source of views presented in §4.)

Strawson, P.F. (1964) 'Identifying Reference and Truth-Values', *Theoria* 30: 96-118.(Source of views presented in §4.)

Strawson, P.F. (1972) *Subject and Predicate in Logic and Grammar*, London: Methuen. (Source of views presented in §4.)

Whitehead, A.N. and Russell, B.A.W. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7, esp. vol. 1, *14. (Source of views presented in §3.)

Discourse semantics

Discourse and its interpretation have interested philosophers since ancient times, and have been studied in different areas of philosophy such as rhetoric, the philosophy of language and the philosophy of literature. Discourse has attracted interest from philosophers working in the continental tradition, and it received considerable attention in the 1980s from analytic philosophers, philosophers of language, linguists, cognitive scientists and computer scientists; within these fields a formal, logical analysis of discourse interpretation, or discourse semantics, has emerged.

Discourse semantics arose in an attempt to solve certain problems that affected formal theories of meaning for single sentences. These problems had to do with the interpretation of pronouns and other 'anaphoric' elements in language. A detailed examination of the data showed that the meaning of an individual sentence in a discourse could depend upon information given by previous sentences in the discourse. To analyse this dependence, discourse semantics developed a formal analysis of a discourse context and of the interaction between the meaning of a sentence and the discourse context in which it is to be interpreted. The essential idea of discourse semantics is that the meaning of a sentence is a relation between contexts. The 'input' discourse context furnishes the information needed to interpret the anaphoric elements in the sentence; the information conveyed by the sentence when added to the input context yields a new, or 'output', discourse context that can serve to interpret the next sentence in the discourse.

1 Static semantics for discourse and its difficulties

To understand the development of discourse semantics, one must look at the problems that motivated its development. To do this, one must know something about the interpretation of the constituents of discourse - the constituent sentences and their constituent phrases as treated in standard formal semantics. In the 1960s, Richard Montague developed a very influential framework, Montague Grammar, for analysing how the meaning of a complex expression depends on the meanings of its constituents (see Compositionality; Possible worlds; Semantics). One might think that there is nothing more to the interpretation of a discourse than simply building up the meanings of its constituent sentences and then combining them. But once a clear and precise proposal for the semantics of sentences was given by Montague Grammar, several problems emerged that showed that the interpretation of discourse would not be so simple. There are two related phenomena in semantics that cause difficulty for Montague Grammar, namely pronominal anaphora and temporal anaphora, and these have spurred advances in the analysis of discourse.

Pronominal anaphora. One problem for the Montague Grammar approach to discourse is the interpretation of anaphoric pronouns. Pronominal anaphora occurs when a pronoun refers back to some word or phrase in the preceding discourse. In the discourses below, the anaphoric pronouns 'he' and 'it' refer back respectively to a 'farmer' and a 'donkey'.

(1) A farmer owns a donkey and he beats it.
(2) A farmer owns a donkey. He beats it.
(3) If a farmer owns a donkey, he beats it.

In semantics the meaning of a discourse like (1) can be given by a formula of some logical language with a clear model-theoretic interpretation - this formula is known as the logical form of the sentence. Montague Grammar uses the language of intensional, higher-order logic, but for our purposes the only thing we need to know about this language is that indefinite noun phrases like 'a farmer' are translated as existential quantifiers and pronouns are translated as bound variables. Thus, a sentence such as (1) has the translation:

$\exists x \exists y (\mathrm{farmer}(x) \ \& \ \mathrm{donkey}(y) \ \& \ \mathrm{owns}(x, \ y) \ \& \ \mathrm{beats}(x, \ y))$

The first problem for Montague Grammar is that since it gives the meaning of a discourse sentence by sentence, it is unclear how to connect the pronouns in the second sentence of (2) with the existential quantifiers that bind them in the first sentence. Another problem emerges with the attempt to treat (3), which is an example of the 'donkey sentence problem' originally discovered by Geach (1963). The meaning of (3) is captured by the formula:

$$\forall x \forall y((\text{farmer}(x) \,\&\, \text{donkey}(y) \,\&\, \text{owns}(x,\,y)) \;\rightarrow\; (\text{beats}(x,\,y))$$

But Montague Grammar has no uniform translation of the indefinite determiner *a* that yields a correct treatment of (1) and (3) satisfactorily; it does not provide any explanation of why the translation of indefinites is context-sensitive in this way, nor does it handle inter-sentential anaphora (see Anaphora).

Temporal structure of discourse. Observations by Barbara Partee and people working on the analysis of French tense in discourse first brought attention to a facet of the meaning of tenses that was missing from the best analyses of tense of the day. Those analyses took tenses to be temporal operators of the sort found in tense logic. In a French discourse like:

(4) Pierre entra dans le salon. Il s'assit sur la banquette. Il s'endorma.
(Pierre entered the room. He sat down on the sofa. He fell asleep.)

the three events introduced (Pierre's entering, his sitting down, and his falling asleep) all occur in the past if (4) is true. This much the analyses of tense were able to capture. But they miss a crucial generalization about texts like (4). The events described there are portrayed as occurring not only in the past but also in a definite sequence - that in which they are introduced in the discourse. We naturally understand the story as telling us that Pierre's entering the room occurred prior to his sitting down which in turn occurred prior to his falling asleep. On the operator view of tenses, a past-tensed sentence of the (logical) form $P\varphi$ is true just if φ holds for some time prior to the moment of speech; this view is incapable of capturing the contextual sensitivity shown in (4). Similarly incapable are views on which verbs are treated as predicates of events and the past tense introduces a relation of 'earlier than' between the event and the moment of speech (see Tense and temporal logic).

2 Dynamic semantics for discourse

Motivated by the problems with the interpretation of pronouns in discourse, Hans Kamp (1981) and Irene Heim (1982) independently proposed not only a solution to the interpretation of anaphoric elements but a new way of looking at discourse semantics. They redefined the semantic contributions in a discourse of a sentence and its constituents. In Montague Grammar, among other accounts of discourse interpretation, the contribution of a sentence is a proposition, or, formally, a set of possible worlds in which the sentence is true. Such a proposition contributes to the content of a discourse in a simple way: the meaning of a discourse is just the intersection of all the propositions that are the meanings of the discourse's constituent sentences. For Kamp and Heim, a sentence *S* when interpreted in a discourse *D* no longer simply yields a set of worlds; *S* may also yield a set of discourse entities or 'discourse referents' to which elements of subsequent discourse may refer. This approach came to dominate semantics in linguistics during the 1980s and 1990s. What follows develops Kamp's approach to discourse semantics, known as 'Discourse Representation Theory' or DRT.

DRT assigns a truth-conditional meaning to a natural language discourse in two steps: the 'DRS construction procedure' and the 'correctness definition'. The first step is to construct a representation of the content of the discourse (known as a 'Discourse Representation Structure' or DRS). The ontological status of a DRS has been subject to some debate: some view it as a partial model of the discourse, others view it merely as an alternative logical form for English, still others as some sort of conceptual structure. Nevertheless, the structure of a DRS is precisely defined. It consists of a pair of sets: (1) a set, or 'universe', of discourse referents, and (2) a set of 'conditions'. Discourse referents are analogous to the domain of a partial model - they are objects talked about in the discourse - while conditions are property ascriptions to these discourse individuals. The technical definition of a DRS and DRS condition is given through a doubly recursive definition. In thinking of DRSs as partial models, DRSs as abstract structures should be distinguished from the language used to describe them. The DRS language uses a box notation - the upper part of the box lists the discourse referents in the universe of the DRS, while the bottom part of the box describes the conditions. Kamp's notation is in fact reminiscent of Peirce's existential graphs. For instance, the DRS for 'a farmer owns a donkey' is given in (K1); it tells us that the discourse speaks of two entities, one a farmer (*x*), one a donkey (*y*), and that, according to the discourse, the farmer owns the donkey.

(K1)

$$\boxed{\begin{array}{c} \underline{\quad x \quad y \quad} \\[4pt] \text{farmer}(x) \\ \text{donkey}(y) \\ \text{owns}(x, y) \end{array}}$$

While Kamp develops a top-down approach to DRS construction, Wada and Asher (1986), Zeevat (1989) and Asher (1993) develop a construction procedure whereby each lexical element contributes some sort of DR-theoretic structure. These then combine following the syntactic structure of the sentence to produce the DRS for the sentence in a manner that is in the tradition of compositional Montague semantics. Noun phrases like 'a farmer' or 'a donkey' introduce a novel discourse referent into the DRS as well as a condition on that discourse referent; verbs introduce conditions on the discourse referents introduced by the noun phrases that are their syntactic arguments. Anaphoric pronouns, following the original Kamp treatment of DRT (see Kamp 1981), are treated as bound variables; they introduce new occurrences of discourse referents that have been introduced by noun phrases in antecedent discourse.

The construction of a DRS for a discourse proceeds incrementally, sentence by sentence, exploiting the syntactic parse of each sentence, and building a DRS for the discourse as a whole from DRSs for each constituent sentence. More precisely, if $\mathbf{K}_m$ is the DRS derived from the first m sentences and K_{m+1} is the DRS derived from the $m + 1$st sentence, then the DRS for the discourse of $m + 1$ sentences is just the DRS that combines the universes of $\mathbf{K_m}$ and K_{m+1} and their condition sets:

$$\langle (\mathrm{U}_{\mathbf{K}_m} \cup \mathrm{U}_{K_{m+1}}), \ (\mathrm{Con}_{\mathbf{K}_m} \cup \mathrm{Con}_{K_{m+1}}) \rangle$$

So in constructing a DRS for (2), for instance, we would add to the DRS in (K1) simply the condition ' $\text{beats}(x, \ y)$' to get a correct representation for (2):

(K2)

$$\boxed{\begin{array}{c} \underline{\quad x \quad y \quad} \\[4pt] \text{farmer}(x) \\ \text{donkey}(y) \\ \text{owns}(x, y) \\ \text{beats}(x, y) \end{array}}$$

Let us now briefly look at DRT's treatment of sentence (3) given in §1. We use the same analysis of indefinite noun phrases and pronouns as introduced above, together with construction procedures for conditionals and universal quantifiers (such as the quantifier expressed by 'every farmer'). Both constructions introduce what is known as a 'complex condition' in a DRS; complex conditions consist of DRSs as arguments to some operator. For instance, 'if…, then… ' introduces the following relation on DRSs, where *P* and *Q* are variables for DRSs.

$$\lambda P \lambda Q \quad \boxed{\begin{array}{c} \underline{\qquad\qquad} \\[4pt] P \quad \Rightarrow \quad Q \end{array}}$$

So when we have two clauses linked by 'if…, then… ', the first clause gives us a DRS bound by *P*, while the second gives us a DRS which replaces *Q*. For example, 'if Mary is happy, then Fred is happy' would yield the following DRS:

$$\boxed{\begin{array}{|c|} \hline \begin{array}{|c|} \hline x \\ \hline \mathrm{Mary}(x) \\ \mathrm{happy}(y) \\ \hline \end{array} \Rightarrow \begin{array}{|c|} \hline y \\ \hline \mathrm{Fred}(y) \\ \mathrm{happy}(y) \\ \hline \end{array} \\ \hline \end{array}}$$
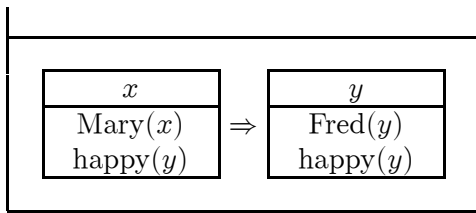
If we choose the right discourse referents for the pronouns 'he' and 'it', we get the right DR-theoretic translation for (3) in a uniform way from the treatment of the indefinite noun phrase and pronouns:

(K3)
$$\boxed{\begin{array}{c} \begin{array}{|c|} \hline x \quad y \\ \hline \mathrm{farmer}(x) \\ \mathrm{donkey}(y) \\ \mathrm{owns}(x,y) \\ \hline \end{array} \Rightarrow \begin{array}{|c|} \hline \quad \\ \hline \mathrm{beats}(x,y) \\ \hline \end{array} \end{array}}$$
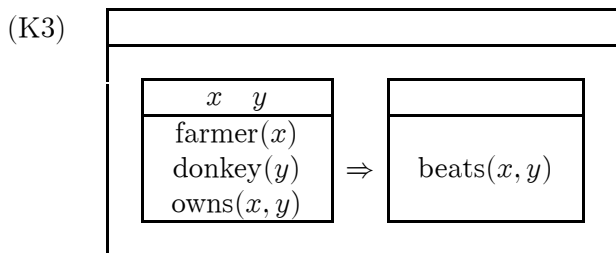
Several forms of complex conditions have been proposed in DRT to handle various connectives and quantifiers. Most prevalent are those of the form $K \Rightarrow K'$ used to represent sentences that contain a conditional or universal quantifier, and those of the form $\neg K$ which result from processing sentences in which a verbal or sentence-level negation occurs.

In order to understand what these complex conditions mean we must look at the second component of DRT, the correctness definition. The correctness definition tells us what conditions must obtain in order for a DRS, thought of as a partial model, to be properly embedded in a Tarskian model of the sort familiar from first-order logic (see Model theory). If the DRS for a discourse can be properly embedded in a model *M*, then we say that the discourse is true in *M*. A 'DRS model' is a pair <*D, I*> where *D* is a non-empty set (a domain of individuals) and *I* is a function that assigns *n*-tuples from D^n to atomic *n*-ary conditions of DRSs. Atomic conditions are those conditions that are derived from natural-language nouns, verbs and some adjectives - the sort that are contained in the DRSs for (1) and (2). A DRS *K* is properly embedded in a DRS model *M* if and only if there is a function *f* from the universe of *K* into the universe of *M* such that all the conditions of *K* are satisfied in *M*. What we have to do now is to define satisfaction of a condition in *K*. Intuitively the idea is simple: for a condition in the DRS to be satisfied in a model, there must be a mapping or embedding of the universe of *K* in the universe of *M* such that the objects $f(x_1),\dots,f(x_n)$ bear the relation that the DRS condition predicates of $x_1,\dots,x_n$. For an atomic DRS condition of the form $\Psi(\alpha_1,\dots,\alpha_n)$, where $\alpha_1,\dots,\alpha_n$ are discourse referents, $\Psi(\alpha_1,\dots,\alpha_n)$ is satisfied in a DRS model *M* relative to an embedding function *f* if and only if $\langle f(\alpha_1),\dots,f(\alpha_n)\rangle$ is an element of the interpretation of Ψ in *M*. A complex condition of the form $K \Rightarrow K'$ is satisfied in a DRS model *M* relative to an embedding function *f* if and only if for every function *g* that extends *f* to a proper embedding of *K* there is an extension *h* of *g* that is a proper embedding of K'. A complex condition of the form $\neg K$ is satisfied in a DRS model *M* relative to an embedding function *f* if and only if there is no function *g* that extends *f* to a proper embedding of *K*. Intuitively, we see how the interpretation of these DRS conditions captures the definition of a material conditional, a universal quantifier and negation. Technically we have a simultaneous recursive definition of proper embedding of a DRS in a model and of the satisfaction of a DRS condition with respect to a model. (For details see Kamp and Reyle 1993.)

In applying this definition of truth to (K3), we get the intuitively right truth-conditions for the sentence: for every pair of objects such that one is a farmer and the other is a donkey and the farmer owns the donkey, the farmer beats the donkey. These truth-conditions are the same as those for the universally quantified sentence, 'every farmer who owns a donkey beats it'. While some problems in the interpretation of pronouns have been uncovered in further studies on anaphora in a dynamic semantic setting, the dynamic semantic approach represents a considerable advance over Montague Grammar's approach to the interpretation of pronouns in discourse.

DRT and temporal anaphora. Following Davidson (1967), DRT takes verbs to introduce conditions on

eventualities (events or states) (see Davidson, D. §2). DRT postulates a complex algorithm that serves to connect by means of temporal relations the eventuality introduced by the main verb of a sentence with the eventualities already present in the DRS built up from previous discourse. In declarative form, we formulate the results of the algorithm as follows. If α and β are two sentences that follow one after another in the discourse and both introduce eventualities with a verb in the *passé simple*, we have (where $<$ is the 'earlier than' relation):

$$\mathrm{PS}(\alpha)\ \&\ \mathrm{PS}(\beta)\ \rightarrow\ e_\alpha\ <\ e_\beta$$

If past tense also implies that the event described by the verb occurs prior to the speech time, this rule in effect predicts just the right temporal interpretation of discourse (4). The complex DRS construction procedure has been applied to basic tense forms of French and English (see Kamp and Reyle 1993 for an extensive discussion and bibliography).

Limitations of dynamic semantics. DRT makes an important contribution to our understanding of how the discourse context affects pronominal and temporal interpretation. But DRT does not do justice to the complex interaction of pragmatic and semantic factors in discourse, and this leads to incorrect predictions about the way anaphora, both temporal and pronominal, is treated in DRT.

A clear example of where DRT goes wrong is with its analysis of temporal anaphora. DRT is right to make the contributions of tense depend upon the discourse context. But it attempts to make the temporal structure of a text almost completely dependent on the tense forms used in the text. In most natural languages, however, the temporal structure of the events introduced in a text is underdetermined by the sequence of tense forms. In particular, the rule above, which is a consequence of the DRS construction procedure, is false for French - a point of which some of the earliest workers on tense in DRT were aware - or English. Consider the following examples (from Lascarides and Asher 1993):

(5a) John entered the room. Fred greeted him.
(5b) John fell. Fred pushed him.

These two discourses employ the same sequence of tense forms, yet they suggest different temporal structures. DR-theorists have been forced to revise the construction procedure and to abandon the view that the tense forms and the order of the sentences in a discourse alone determine temporal structure. This conclusion follows not only from an examination of the English simple past but also from a careful look at the data concerning the French *plus que parfait* (see Asher and Bras 1992) and the English pluperfect (Lascarides and Asher 1992).

One might ask, what in combination with tense sequences determines the temporal structure? One proposal is that a more developed view of discourse structure is needed. Originally suggested by the computer scientist and artificial intelligence (AI) researcher Jerry Hobbs (1985), this thesis has been worked out in the context of a formal discourse semantics by Lascarides and Asher (1993). This approach to tense is part of a more general view: discourse structure has systematic effects on discourse interpretation, which it is the task of discourse semantics to study. As we shall see, both temporal and pronominal anaphora require a more elaborate discourse semantics.

3 More elaborate discourse semantics

When interpreters read a text, they naturally understand it as divided into parts related in various ways. Some parts form a narrative, others furnish a background to a narrative or an explanation of some action, still others may elaborate other parts of the text. Sometimes these divisions and the relations between them are signalled by punctuation and the presence of certain words in the text. But often interpreters infer the appropriate divisions and discourse relations without these cues. This sort of structure is completely missing from the treatment of discourse in DRT, but it has been recognized as an important aspect of discourse in other disciplines - for example, literary theory and the natural language processing (NLP) area of AI (see Artificial intelligence). In the mid-1970s, computer scientists like Roger Shank claimed text-understanding systems and text-information retrieval systems to be one of the central goals of NLP research in AI. This early interest led to a substantial amount of research on discourse structure in the mid to late 1980s (Polanyi 1988; Hobbs 1985; Grosz and Sidner 1985). This work has concentrated on the discourse structure of written texts, though analyses of highly structured, task-oriented dialogues have yielded important insights. Many divergent positions have been taken, but several uncontroversial conclusions have been drawn from these studies: (1) discourse is structured and can be represented by a set of
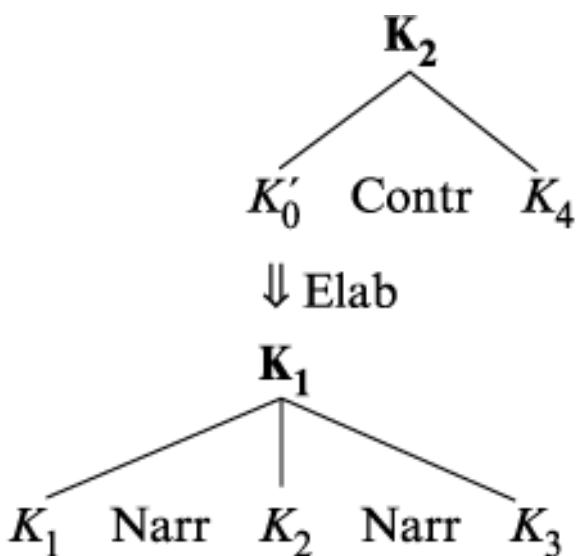
objects (propositions or bits of text) related by discourse relations; (2) this structure is hierarchical and recursive. For discourse semantics, we must integrate the view of discourse structure elaborated by these authors with a theory of meaning. This connection between semantics and discourse structure, however, is lacking in most of the discourse theories in AI.
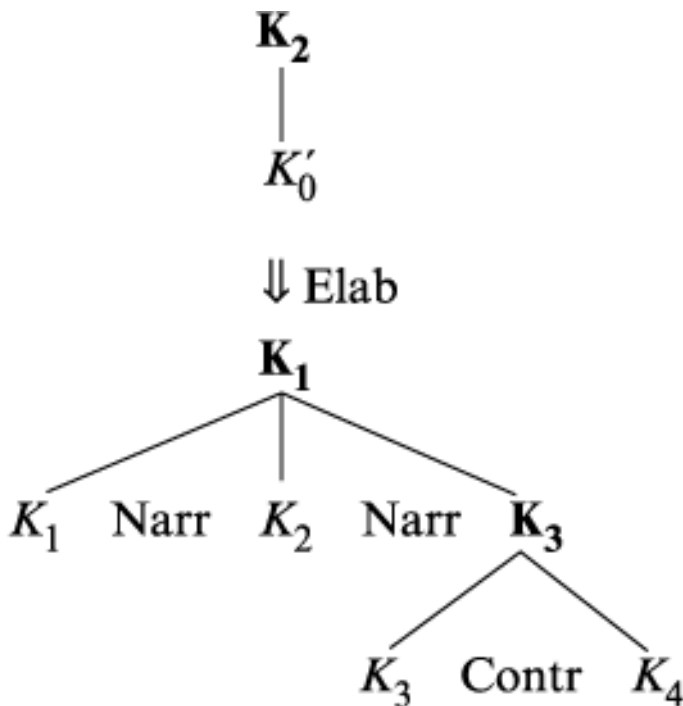
A richer notion of discourse structure is incorporated into the framework of dynamic semantics in Asher (1993), resulting in a nontrivial extension of DRT known as 'segmented Discourse Representation Theory' or SDRT. SDRT is one way to integrate discourse structure for texts and semantics. Another approach can be found in the works of Scha and his colleagues, but it exploits a different version of dynamic semantics.

According to SDRT, every discourse is analysed into a representational structure called a 'segmented DRS' or SDRS, and this is in turn embedded into a complete model that furnishes truth-conditions for the discourse. The model theory for SDRSs is somewhat involved, because the models must allow for quantification over and reference to dynamic propositions (for details see Asher 1993 or Asher 1996). An SDRS, like a DRS, is a pair of sets; but unlike a DRS the elements of the first set are either DRSs or SDRSs, while the second pair consists of a set of conditions in which discourse relations are predicated of elements in the first set. In comparing an SDRS for a narrative text with a DRS for the same text, we would see that the DRS is segmented into elements of the first set in the SDRS with discourse relations holding between these elements. These elements or 'constituents' of an SDRS serve several useful functions. First of all, they constitute referents for anaphoric pronouns (like 'this' in the example below) whose antecedents are facts or propositions:

(6) One plaintiff was passed over for promotion. Another didn't get a pay increase for five years. A third received a lower wage than men doing the same work. But the jury didn't believe this (any of this).

In the example above, we appear to have two possible antecedents (depending to some degree on how this is stressed): the proposition formed by the discourse as a whole and the proposition given by the last sentence. In SDRT these two anaphoric possibilities and only these two anaphoric possibilities are predicted; they result from two different, acceptable places to which the new information contained in the last sentence can be attached to the SDRS built up from the previous three sentences. Below we have a pictorial representation of the two SDRSs that could be constructed for (6). $K_1, \ldots, K_4$ represent the DRSs derived from the four sentences of the discourse; K_0' is a topic that results from the SDRS construction and the constraints on Narration (every Narration must have a nontrivial topic); the shadowed letters represent complex constituents or constituents that are SDRSs. The lines represent membership, while $\Downarrow$ represents a particular type of discourse relation (like Elaboration) that introduces a hierarchical structure into an SDRS. Narr stands for Narration, Elab for Elaboration, and Contr for Contrast. Notice the two different positions of attachment for K_4.

$$
\begin{array}{c}
\mathbf{K_2} \\
\swarrow \qquad \searrow \\
K_0' \quad \text{Contr} \quad K_4 \\[1em]
\Downarrow \text{Elab} \\[0.5em]
\mathbf{K_1} \\
\swarrow \quad | \quad \searrow \\
K_1 \quad \text{Narr} \quad K_2 \quad \text{Narr} \quad K_3
\end{array}
$$

$$\mathbf{K_2}$$
$$|$$
$$K_0'$$

$$\Downarrow \text{Elab}$$

$$\mathbf{K_1}$$

K_1 Narr K_2 Narr $\mathbf{K_3}$

K_3 Contr K_4

The crucial task for a discourse semantics like SDRT is to specify a procedure that will enable us to construct a discourse representation. In DRT the discourse representation is built up from the syntactic structure and a translation procedure for syntactic elements in the parse tree. SDRT exploits the DRS construction procedure to build DRSs for clauses, and it also makes use of the constraints DRT places on anaphora resolution. But in SDRT we need more information to be able to deduce where to attach new information in the antecedently built-up representation and by means of which discourse relation to attach it. The constraints on appropriate attachment sites are defined configurationally, following much of the work on discourse structure in AI. But the difficult part is to infer the appropriate discourse relation. Sometimes certain cue words or phrases indicate the presence of particular discourse relations, but those discourse relations may be present even when there are no cues. For instance, consider the following pairs:

(7a) John went to jail. He embezzled funds from the pension account.
(7b) John went to jail, because he embezzled funds from the pension account.

(8a) Mary went to school. She met with her students. She went home.
(8b) Mary went to school. Then she met with her students. Afterwards she went home.

Discourses (7a) and (7b) on the one hand and (8a) and (8b) on the other, lend themselves to the same interpretation. In the examples in (7), the second clause or sentence gives us the explanation for what happened in the first; while in (8), we read the clauses in each as narrating a sequence of events that are in the order that they are introduced in the text. Yet only in (7b) and (8b) are these interpretations fully determined by the semantic contributions of special 'clue' words like 'because' or the temporal adverbials in (8b). In (7a) and (8a), we infer the presence of an Explanation relation or a Narration relation based on pragmatic principles and the contents of the constituents the discourse relations tie together in the SDRSs for these discourses. However, this inference is not deductively valid; within a richer discourse context, we can cancel, for instance, the inference that the second sentence in (7a) offers an explanation of what happened in the first:

(7c) John went to jail. He embezzled funds from the pension account. But that is not why he went to jail.

But it also seems that the explanation relation is undeniably there in (7a). Thus, we must have a means of inferring discourse relations by default; the default or 'non-monotonic' inference of a discourse relation R can be overridden in cases where we have information in the context that is inconsistent with our conclusion that R holds. SDRT uses facts about the discourse context, the contents of the constituents including the lexical contents of particular words, world knowledge and pragmatic principles, to infer discourse relations non-monotonically. An attempt to code all

of this knowledge explicitly for certain simple examples is given in Lascarides and Asher (1993).

Once we have inferred that some discourse relation *R* relates new information to some available attachment point in the SDRS constructed from the preceding discourse, we can update this SDRS with the new information. But this updating does not consist in simple addition. It also requires in many cases the revision of the contents of the constituents in the SDRS. It is in the process of constituent revision that anaphora resolution and the determination of temporal structure are accomplished in SDRT. For instance, consider again the discourse in (4). The *passé simple* of the verbs in this text all introduce the simple DRS condition saying that the eventuality described by the verb occurred prior to the speech point. There is no attempt in the semantics to link up the eventuality with the temporal structure built up antecedently, as there is in DRT. But now the underlying logical inference mechanisms of SDRT discussed above lead us to conclude that Narration relates the constituents α and β derived from the first two sentences and the constituents β and χ derived from the second and third sentence. According to SDRT, Narration has certain temporal effects: $\mathrm{Narration}(\alpha, \beta) \to e_\alpha < e_\beta$. So once we have attached the three constituents in (4) with Narration, we immediately conclude that the text gives us a sequence of events that follows the order in which these events are introduced in the text. The effect of Narration is one way to formulate precisely part of Grice's maxim 'be orderly'. The order of presentation of the events in the text reflects their temporal order (see Implicature §4). But the pragmatic principles appealed to in the formalization of Lascarides and Asher (1993) and used to infer that Narration holds between these constituents are a good deal more specific than the general Gricean maxims of conversation, and they are expressed in logic.

In (5b) additional information contained in the two clauses leads us to defeat the inference to Narration and to conclude that a different discourse relation - Explanation - holds between the two constituents of the SDRS. Explanation has different temporal consequences from Narration: $\mathrm{Explanation}(\alpha, \beta) \to \neg(e_\alpha < e_\beta)$ and, where ' $>$' is a weak conditional used to represent default patterns, $\mathrm{Explanation}(\alpha, \beta) > (e_\beta < e_\alpha)$. So we conclude that in (5b) the pushing came *before* the falling, as intuitions would dictate. Thus, while DRT attempted to derive the temporal structure of a text from the semantics of the sequence of tenses in a discourse alone, SDRT exploits not only the compositional semantics but also many other sources of information that contribute to discourse structure to determine temporal structure.

4 Conclusions

Because of the rapidity with which the field of discourse semantics has evolved, it is hard to say exactly where the field will develop or even whether it will continue as a field. It is quite unclear whether DRT or SDRT will remain active research areas. SDRT has been applied not only to the central discourse semantic concerns of temporal and pronominal anaphora, but it has also helped in analysing the spatiotemporal structure of texts (Asher *et al.* 1994), the role of discourse structure in lexical disambiguation (Asher and Lascarides 1995), and the interactions between discourse structure and speakers' and interpreters' cognitive states - in particular their beliefs and discourse goals (Asher and Lascarides 1994). To advance further, SDRT and other formalisms for discourse semantics require a better understanding of the organization of the lexicon and of the role of general world knowledge. As discourse semantics moves to give a formally precise analysis (capitalizing on empirical studies) of discourse interpretation in dialogue, other factors like the interaction between semantics and speech acts will have to be taken into account (see Speech acts).

See also: Pragmatics

NICHOLAS ASHER

References and further reading

Asher, N. (1993) *Reference to Abstract Objects in Discourse*, Dordrecht: Kluwer.(This book (referred to in §3) introduces SDRT to take account of rhetorical relations and their effects on truth-conditional content. It applies SDRT to propositional anaphora and VP ellipsis.)

Asher, N. (1996) 'Mathematical Treatments of Discourse Contexts', *Proceedings of the Tenth Amsterdam Conference on Formal Semantics*, Amsterdam: ILLC Publications, vol. 1, 21-40.(This article gives a revised, improved version of SDRT. It is a much shorter introduction to SDRT than Asher 1993.)

Asher, N., Aurnague, M., Bras, M., Sablayrolles, P. and Vieu, L. (1994) 'Computing the Spatiotemporal Structure of Discourse', in H. Hunt, R. Muskens and G. Rentier (eds) *Proceedings of the First International Workshop on Computational Semantics*, Tilburg: Tilburg University Publications, 11-20.(This article uses

SDRT to model the spatio-temporal structure of discourse.)

Asher, N. and Bras, M. (1992) 'The Temporal Structure of French Texts within Segmented Discourse Representation Theory', in M. Aurnague, A. Borillo and M. Bras (eds) *Semantics of Time, Space, Movement and Spatio-Temporal Reasoning*, Working Papers of the 4th International Workshop on Time and Space, IRIT Working Papers, Toulouse: Université Paul Sabatier, 203-18.

Asher, N. and Lascarides, A. (1994) 'Intentions and Information in Discourse', *Annual Proceedings of the American Computational Linguistics Meetings*, 34-41.(This paper addresses the information flow between discourse structure and cognitive states in a couple of examples. Referred to in §4.)

Asher, N. and Lascarides, A. (1995) 'Lexical Disambiguation in a Discourse Context', *Journal of Semantics* 12 (1): 69-108.(This paper applies SDRT to lexical disambiguation and shows how discourse structure can affect this task. Referred to in §4.)

Davidson, D. (1967) 'The Logical Form of Action Sentences', in N. Rescher (ed.) *The Logic of Decision and Action*, Pittsburgh, PA: The University Press, 81-95.(In this article, Davidson proposes that verbs are predicates of events.)

Geach, P.T. (1963) *Reference and Generality*, Ithaca, NY: Cornell University Press.(This book is a collection of Geach's influential articles, in which among other things the puzzle about donkey sentences (referred to in §1) is introduced.)

Grice, H.P. (1967) *Logic and Conversation. William James Lectures*, repr. in *Studies in the Ways of Words*, Cambridge, MA: Harvard University Press, 1989.(The introduction of conversational 'maxims', which are referred to in §3.)

Groenendijk, J. and Stokhof, M. (1990) 'Dynamic Predicate Logic', in H. Kamp (ed.) *Dyana Deliverable*, 2.5a. (This paper introduces an alternative approach to dynamic semantics from DRT as presented in Kamp 1981.)

Grosz, B. and Sidner, C. (1985) 'The Structure of Discourse Structure', Palo Alto: SRI, SRI Technical Note 369. (This paper gives a very influential view of discourse structure and argues for a parallel between discourse structure and the intentions of the speaker. Referred to in §3.)

Heim, I. (1982) *The Semantics of Indefinite and Definite Noun Phrases*, Amherst, MA: University of Massachusetts Press.(Referred to in §2; Heim's Ph.D. thesis offers an alternative way of developing dynamic semantics to Kamp 1981 and is close to the views presented in Groenendijk and Stokhof 1990.)

Hobbs, J. (1985) 'On the Coherence and Structure of Discourse', Stanford, CA: Center for the Study of Language and Information, report no. CSLI-85-37.(This paper introduces and argues for the idea that various anaphoric phenomena may be dependent upon discourse structure. Referred to in §§2-3.)

Kamp, H. (1981) 'A Theory of Truth and Semantic Representation', in J. Groenendijk, T. Janssen and M. Stokhof (eds) *Formal Methods in the Study of Language*, Amsterdam: Mathematisch Centrum Tracts, 277-322; repr. in J. Groenendijk *et al.* (eds) *Truth, Interpretation and Information*, Dordrecht: Foris, 1984, 1-41.(This article (referred to in §2) contains the original presentation of DRT. It is difficult to read but is the principal source for many of the developments in dynamic semantics.)

Kamp, H. and Reyle, U. (1993) *From Discourse to Logic; Introduction to Model-Theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Dordrecht: Kluwer.(This book is an introductory textbook to DRT that extends considerably the semantic coverage of the theory from that given in Kamp 1981. Referred to in §2.)

Lascarides, A. and Asher, N. (1992) 'The Pluperfect in Narrative Discourse', in M. Aurnague, A. Borillo, M. Borillo and M. Bras (eds) *Semantics of Time, Space, Movement and Spatio-Temporal Reasoning*, Working Papers of the 4th International Workshop on Time and Space, IRIT Working Papers, Toulouse: Université Paul Sabatier, 183-217.(A condensed version of this paper called 'A Semantics and Pragmatics of the Pluperfect' appears on pages 250-9 of *Proceedings of the 6th Conference of the European Chapter of Computational Linguistics*, published in Utrecht in 1993. These papers discuss the pluperfect in narrative discourse from the perspective of SDRT.)

Lascarides, A. and Asher, N. (1993) 'Temporal Interpretation, Discourse Relations, and Commonsense Entailment', *Linguistics and Philosophy* 16: 437-93.(This paper criticizes the analysis of tense within DRT and proposes a new simplified account using the SDRT framework. Referred to in §§2 and 3.)

Partee, B. (1973) 'Some Structural Analogies between Tenses and Pronouns in English', *Journal of Philosophy* 70: 601-9.(In this article Partee argues that tenses like the simple past exhibit a contextual sensitivity in interpretation similar to that of anaphoric pronouns.)

Polanyi, L. (1988) 'A Formal Model of the Structure of Discourse', *Pragmatics* 1-88.(This article contains another influential approach to discourse structure, with powerful treatments of interruptions and discourse repairs. Referred to in §3.)

Prust, J. and Scha, R. (1990) 'A Discourse Perspective on Verb Phrase Anaphora', in M. Stokhof and L. Torenvliet (eds) *Proceedings of the Seventh Amsterdam Colloquium*, Amsterdam: ITLI Publications, 451-74. (This article develops the theory of discourse parallelism in detail and points to the use of dynamic semantics as developed by Groendijk and Stokhof as the interpretive component.)

Scha, R. and Polanyi, L. (1988) 'An Augmented Context Free Grammar for Discourse', *Proceedings of COLING*, Budapest.(This article proposes a 'discourse syntax', in which trees model discourse structures, together with an interpretation of the leaves of those trees. The interpretation uses Montague Grammar, however, and is not dynamic.)

Wada, H. and Asher, N. (1986) 'BLDRS: A Prolog Implementation of LFG and DR Theory', *Proceedings of the 11th International Conference on Computational Linguistics*, Bonn, 540-5.(This was one early implementation of DRT and suggested a compositional treatment of dynamic semantics.)

Zeevat, H. (1989) 'A Compositional Approach to Discourse Representations', *Linguistics and Philosophy* 12: 95-131.(This paper gives a detailed treatment of how to build up DRSs in a compositional way, similar to the construction of formulas representing meanings of expressions in Montague Grammar. Referred to in §2.)

Emotive meaning

Emotive meaning contrasts with descriptive meaning. Terms have descriptive meaning if they do the job of stating facts: they have emotive meaning if they do the job of expressing the speaker's emotions or attitudes, or exciting emotions or attitudes in others. Emotivism, the theory that moral *terms have only or primarily emotive meaning, is an important position in twentieth-century ethics. The most important problem for the idea of emotive meaning is that emotive meaning may not really be a kind of meaning: the jobs of moral terms supposed to constitute emotive meaning may really be performed by speakers using moral terms, on only some of the occasions on which they use them.*

There are two components in emotivist accounts of the function and meaning of moral terms. One is a matter of relations to the speaker: moral assertions serve to express the speaker's emotions or attitudes. The other is a matter of relations to the audience: moral assertions serve to commend things, or to arouse emotions or attitudes in the audience.

The most celebrated accounts of emotive meaning were developed by A.J. Ayer and C.L. Stevenson. Ayer (1936) argued on general metaphysical and epistemological grounds that moral terms can only express and excite emotions. Stevenson (1937) developed a more detailed theory, relying more on distinctively ethical considerations (see Emotivism). These theories are liable to make moral discussion seem irrational, and to make no distinction between moral argument and propaganda. R.M. Hare developed a theory designed to remedy these defects. He argued that sentences using paradigm moral terms like 'good', 'right' and 'ought' are really disguised imperatives. Since there is a logic of imperatives, there is room for rational moral argumentation, and moral argument can be distinguished from propaganda, even though moral assertions do not primarily state facts (see Prescriptivism).

The most important difficulties for the idea of emotive meaning can be raised by asking whether the emotive meaning of moral terms is a matter of the speech act performed by someone using these terms, and, if so, what kind of speech act? One can distinguish between locutionary, illocutionary, and perlocutionary acts (or force). Locutionary acts are simply a matter of uttering certain words with certain senses and referents. Illocutionary acts are done *in* saying things; what illocutionary act one performs in uttering a sentence is determined together by the senses and referents of the words in the sentence and the context. Perlocutionary acts are done *by* saying things. That one performs a certain perlocutionary act is not guaranteed by performing an appropriate illocutionary act: it depends on further variable features of the context. Suppose Bob utters the words 'Down with the aristocrats!' before a large crowd in a revolutionary situation. Bob performs a locutionary act just by saying words with that meaning. In this context, Bob also performs the illocutionary act of *inciting* revolution. Bob's words may also, in that context, have the perlocutionary effect of *provoking* revolution; but this is not guaranteed by the locutionary or the illocutionary force of Bob's utterance. If Bob is sufficiently unpopular, his advocating revolution may actually dampen revolutionary enthusiasm. This classification suggests that the order of explanation typically goes from the locutionary to the illocutionary and the perlocutionary: it is in virtue of the sense and reference of the words one utters and the context that one performs a certain illocutionary act, and that one performs certain perlocutionary acts. The presumption is that an account of meaning will begin with sense and reference; if it begins with illocutionary or perlocutionary acts, it may begin in the wrong place (see Speech acts).

Some early accounts of emotive meaning seem to identify the meaning of moral assertions with perlocutionary acts (for example, arousing emotions). But this seems both to be the wrong place to begin and to raise a further concern. If we think of meaning as a matter of convention, not (mere) causal variation, then causal correlations between utterances and the production of certain effects are not really meanings. Perhaps then emotive meaning is a matter of illocutionary force. Unfortunately, seeing emotive meaning as illocutionary force is also problematic; in addition to the general problem that an account of meaning apparently should not begin with the illocutionary, there is the specific problem that moral terms do not seem *always* to have the right kinds of illocutionary force. Consider the suggestion that 'good' is used to commend. While it may be true that the term 'good' as used in 'This is a good tennis racket', uttered in a sports shop has the illocutionary force of commending a tennis racket, 'good' as used in 'If you can't get a good one there, try the shop down the street', does not obviously seem to commend anything.

Moral arguments raise especially acutely a version of the same problem for the view that emotive meaning is illocutionary force. Consider the argument: 'Telling the truth is good; if telling the truth is good, getting your little brother to tell the truth is good; so getting your little brother to tell the truth is good.' This argument looks valid. But while 'telling the truth is good' in its occurrence as the first premise commends telling the truth, in its occurrence in the second premise it does not seem to commend anything. So, the suggestion is, emotivists cannot account for the validity of some moral arguments, because if meaning is understood as illocutionary force, it is not the same between different occurences of the same words in arguments. This problem is often called 'the problem of unasserted contexts' (in the second premise, 'telling the truth is good' occurs unasserted).

Fans of emotive meaning can respond to these criticisms. Hare (1952), Blackburn (1984) and Gibbard (1990) have all offered solutions to the problem of unasserted contexts. It is certainly true that the meanings of *some* words (like 'promise') seem well explained by explaining the illocutionary act one performs in using them. Moreover, it is not an accident that, for instance, 'good' is often used to commend, while it is an accident if 'fast' is used to commend. Still, an attractive alternative to the theory that 'good' has emotive meaning is that 'good' *means* something like 'meets the relevant standards'. It is by virtue of this meaning that 'good' is often used to commend. Emotive force is then explained by meaning, not vice versa.

See also: Emotivism; Prescriptivism

<div align="right">DAVID PHILLIPS</div>

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, New York: Dover.(Chapter 6 is Ayer's classic argument for and presentation of emotivism.)

Alston, W. (1967) 'Emotive Meaning', in P. Edwards (ed.) *The Encyclopedia of Philosophy*, New York: Macmillan and the Free Press, vol. 2, 486-93.(A trenchant critique of the idea of emotive meaning.)

Blackburn, S. (1984) *Spreading The Word*, Oxford: Oxford University Press.(Chapters 5 and 6 defend a theory giving moral words emotive meaning.)

Gibbard, A. (1990) *Wise Choices, Apt Feelings*, Cambridge, MA: Harvard University Press.(A sophisticated and wide ranging recent defence of, *inter alia*, a kind of emotive meaning.)

Hare, R.M. (1952) *The Language of Morals*, Oxford: Oxford University Press.(Classic defence of the view that moral terms are disguised imperatives.)

Stevenson, C.L. (1937) 'The Emotive Meaning of Ethical Terms', *Mind* 46: 14-31.(Stevenson's initial presentation of emotivism.)

Stevenson, C.L. (1944) *Ethics and Language*, New Haven, CT: Yale University Press.(Book-length development of Stevenson's view.)

Urmson, J.O. (1968) *The Emotive Theory of Ethics*, London: Hutchinson & Co.(Balanced consideration and development of emotive theory.)

Ziff, P. (1960) *Semantic Analysis*, Ithaca, NY: Cornell University Press.(Chapter 6 is an early and trenchant development of the kind of critique of emotive meaning given above.)

Fiction, semantics of

Taken at face value, 'Anna Karenina is a woman' seems true. By using Tolstoi's name 'Anna Karenina' and the predicate 'is a woman' we appear to refer to the character Anna and to attribute to her a property which she has. Yet how can this be? There is no actual woman to whom the name refers. Such problems of reference, predication and truth also arise in connection with representational art and with beliefs and other attitudes.

Meinong distinguishes the 'being' of objects (including fictional objects) from the 'existence' of actual objects such as Napoleon. 'Anna Karenina' refers to a concrete, particular, nonexistent object that has the property of womanhood. However, Meinong's distinction seems to many ontologically suspect. Perhaps, then, being is existence and 'Anna Karenina is a woman' is actually false because 'Anna Karenina' has no referent. Russell in 'On Denoting' (1905) agrees. But how can we explain the apparent contrast in truth between this sentence and the unquestionably erroneous 'Anna Karenina is from Moscow'? Or is it that being is existence but 'Anna Karenina' refers to an abstract, not a concrete, thing - an existent, abstract thing that does not have the property of being a woman but has merely the property of being said, by Tolstoi's novel, to be a woman? Then, however, the meaning of our sentence about Anna no longer parallels that of 'Emily Dickinson is a woman'. Perhaps, as many argue, we only pretend that 'Anna Karenina' refers and that the sentence is true. This position may not adequately explain the intuitions that support Anna Karenina as a genuine object of reference and predication, however.

1 Meinongian and abstract-object views

Meinong distinguishes the 'being' of objects (including fictional objects) from the 'existence' of actual objects such as Napoleon. Although Meinong's position at first blush respects our face-value understanding of sentences about fiction, it encounters difficulties (see Meinong, A.). Fictional characters normally are incomplete: it is not true, according to the novel, that Anna is right-handed; and it is not true, according to the novel, that Anna is not right-handed. Yet how, without violating the law of excluded middle, can there be such a thing? Again, some fictions are inconsistent: in a story, a wood sprite squares the circle. Does the realm of being then include an object that infringes the law of contradiction?

Contemporary Meinongians suggest answers to such difficulties. Thus Parsons (1980) distinguishes sentence negation ('x is not P') from predicate negation ('x is non-P'). He correlates existent objects one to one with the sets of ordinary properties that they have and extends the correlation by introducing nonexistent objects as entities that are correlated one to one with the sets of ordinary properties that are not correlated with existent objects. No object is contradictory, that is, no object satisfies a formula of the form 'x is F and not (x is F)'. But objects may have impossible properties - for example, being P-and-non-P - without implying the truth of any contradictions. (Such objects are impossible objects and cannot exist.) Objects also may be incomplete without violating excluded middle: the lack both of being-P and of being-non-P does not contradict the fact that an object either is P or is not P.

Such theories show that Meinongian views can be defended with far more plausibility than has been thought by philosophers heavily influenced by Russell's and others' identification of being and existence. Nevertheless, these views do not easily capture all aspects of sentences relating to fiction. For one thing, there are problems specific to particular Meinongian approaches. Thus in Parsons' theory, as in non-Meinongian theories that regard characters as collections of properties (for example, Wolterstorff 1980), it is difficult to allow that a novel may introduce two distinct characters assigned the same properties. Again, our epistemic situation and that of Sherlock Holmes are relevantly similar: Holmes senses and reasons about objects in the world, and so do we. How then can we know (as we do) that we exist, given that he - a nonexistent thing - cannot know that he exists? For another thing, the 'face-value' understanding just sketched of sentences relating to fiction - in which in saying 'Anna Karenina is a woman' we state a truth, refer to the character Anna, and attribute to her the property of womanhood - is over simple.

Thus although 'Anna Karenina is a woman' may be true *in the story*, so also is 'Anna Karenina exists'; and if we grant that the same character may occur in different stories and be described contradictorily by them, then it may be true that the character both is P (in one story) and is not P (in another story). On pain of contradiction, Meinongians cannot understand such claims as together implying that objects both have and do not have the

relevant properties. Complex Meinongian treatments of these claims can be devised. But it seems simpler to allow story-relative properties and then take fictional names to designate abstract, but existent, things that have properties such as being-*P*-according-to-story-*s* and being-not-*P*-according-to-story-*t* without having the non-relative properties of being-*P* and of being-not-*P*. We can thus proceed without having to introduce mysterious Meinongian nonexistent, particular things at all. As discussed in §§2-3 below, there are still further alternatives to considering claims relating to fiction as simple, literal truths in which objects of reference are assigned properties. Many writers would say that unless all these alternatives fail, it is unreasonable to commit ourselves to the perplexities of the Meinongian approach.

The one alternative that is clearly unsatisfactory is the unsupplemented Russellian view that a claim such as 'Anna Karenina is a woman' is false (or lacks truth-value, as Fregeans hold). Perhaps such a claim, taken on its own, is indeed false (or expresses no proposition at all, given direct-reference theories of proper names). But such a view leaves unexplained the contrast between that claim and 'Anna Karenina is from Moscow'. Nor does it explain our inclination to ascribe truth to further claims such as 'Sherlock Holmes is a fictional detective', 'Odysseus is the same as Ulysses' and 'Sherlock Holmes is far more famous than any real detective' (Howell 1979; Parsons 1980; van Inwagen 1977). Unsupplemented Russellian views suggest nothing about the semantic or pragmatic factors which may enter into the understanding and acceptance of these further claims. (With qualifications, a similar point also holds for proposals such as that made by Woods (1974) to understand the truth of claims about fiction by appeal to the specific words or sentences that actually occur in the fictions themselves.)

The abstract-existent-object position allows for the further claims just noted, and that fact together with its non-Meinongian way of accepting genuine fictional entities makes it attractive. In van Inwagen's version of this position, characters of fiction are theoretical entities of literary criticism. Sentences in a work of fiction that describe a character must be distinguished from our own sentences about that character. As part of the novel, Tolstoi's sentence 'Anna took the tray' is not about anything, expresses no proposition, is not used by Tolstoi as a vehicle of an assertion. But my description of the novel, using the same sentence, expresses a true proposition about Tolstoi's character: the proposition that the property of taking a tray is ascribed to that character in *Anna Karenina*. Like all entities, that character exists and obeys the laws of excluded middle and of contradiction. The evidence for the existence of characters is provided by sentences such as 'Anna is the principal character of *Anna Karenina*' or 'There is a character in a nineteenth-century novel who is presented in greater physical detail than are any characters in eighteenth-century novels'. We accept these sentences as true, and it seems impossible to paraphrase them in ways that say the same thing yet involve no quantification over characters. Yet these and other sentences can be straightforwardly represented using van Inwagen's apparatus: 'Anna Karenina has the property of being the principal character in *Anna Karenina*'; or (roughly) 'There is a character *c* and a nineteenth-century novel *n* such that there are properties ascribed to *c* in *n* and some of those properties imply more physical detail than any properties ascribed to any characters in any eighteenth-century novels'.

The abstract-existent-object theory provides the most plausible account so far that recognizes genuine fictional objects without accepting Meinongianism. But the theory has never been presented in full detail. Moreover, as used by us the name 'Anna Karenina' (unlike 'Emily Dickinson') is not the usual proper name that, according to direct-reference theories of proper names, refers directly to its bearer through an initial act of baptism. Rather, we refer to the theoretical entity via the name ascribed to it in the novel. As noted, we also attribute to this entity the property of having the property of womanhood ascribed to it in the novel, rather than the simple property of womanhood. Yet the theory by itself offers no particularly plausible explanation why, when we are really treating Tolstoi's character in this way, we none the less use language appropriate to the reference and predication that concern ordinary concrete particulars ('Anna Karenina is a woman').

2 Pretence and make-believe

Instead of directly attacking questions about fictional reference, predication, truth and ontology, perhaps we should focus on the fictional use of language. Clearly (to ignore issues about fictional names), sentences in fiction or about characters in fiction are not distinguished from ordinary literal sentences by special lexical meanings or grammatical constructions. Like ordinary literal sentences, such sentences also may be true (as are various of Tolstoi's claims about Napoleon in *War and Peace*) and may be vehicles of assertion (as, again, are various of Tolstoi's claims). Perhaps, however, as some speech-act theorists have suggested, fictional claims are specially

marked out by their use in communication, for example, by possessing a special illocutionary force (see Speech acts §1).

In one version of this suggestion, in producing fiction the author performs no new illocutionary act analogous to ordinary acts of asserting, questioning or requesting. Instead, authors pretend, in a non-deceptive way, to perform such acts; for example, pretend to be making assertions about events of which they have knowledge. This suggestion fails, for not all pretences to assert produce fiction (for example, putting on an illustrative performance of making a silly claim), and not all productions of fictional sentences are pretences to assert (typing out one's story and mailing it to a publisher).

In another version of the suggestion, the author performs a special new sort of illocutionary act. For example (as Currie (1990) proposes), the author produces sentences with the intention that the audience will *make believe* the content of those sentences (that there is someone named 'Anna Karenina' who is a woman, and so on) through the audience's recognition of the author's intention. Like other pretence and make-believe approaches (including Walton's, described below, which has influenced Currie), this suggestion explains why both Tolstoi and we produce the sentence 'Anna Karenina is a woman' even while we recognize that 'Anna Karenina' denotes no actual person. We do so because that sentence gives the content that we are to make believe is true. (Because the sentence mimics claims describing real persons, we can also easily use it in making believe that there is a real person whom it truly describes.) One might object that the notion of fiction includes nonlinguistic items such as representational paintings and sculptures as well as sentences. However, its defenders argue that the preceding suggestion can be generalized to cover such items.

A deeper objection derives from Walton's nonlinguistic theory of fiction. For Walton (1990), individuals and groups play games of make-believe, the rules of which require the participants to imagine various matters to be true. Props generate fictional truths in such games. (Thus in a game in which the rule is that stumps are to be imagined to be bears, a stump by its presence generates the fictional truth that a bear is present.) Fictions - including linguistic fictions, representational paintings and sculptures - serve as props in games of make-believe. A text constitutes a fiction when, roughly, there is a rule in force that we are to make believe that there are objects such that the words of that text refer to and describe those objects. (Thus we make believe that 'Anna Karenina' is a genuine proper name that directly refers to a Russian woman of whom the story is telling us, and so on.) In Walton's account, the fact that words constitute a fiction depends on their thus being a prop in a game of make-believe, not on the fact that an author intends the audience to make believe that the propositional content expressed by the words is true. For Walton, fiction has nothing special to do with communicative acts of intentional agents; naturally occurring cracks in a rock can spell out a story.

These last points are controversial. (Perhaps the cracks are only treated by us as fiction, without really being so, as Currie argues.) But the basic idea, developed in detail by Walton, that our claims about fiction are to be understood in terms of make-believe provides a deep, powerful explanation of much that concerns those claims. Quite possibly, and as both Walton's and Currie's views suggest, we only make believe that through Tolstoi's sentences about Anna Karenina we get genuine reference, predication and truth.

3 Further questions

Nevertheless, questions remain: about the content of our make-believe, about non-make-believe claims concerning fiction, and about ontology. If we make believe that 'Anna Karenina is a woman' is true, what exact proposition do we make believe is true? If in a piece of literary criticism I assert that Anna is from St Petersburg - or if I compare her with real persons or with other characters ('Anna Karenina is a stronger, more developed personality than is Theodor Fontane's Effi Briest') - then I seem to be stating a sober, literal truth rather than simply making believe that something is true. If you and I independently remark that Tolstoi's character Anna is from St Petersburg, our remarks seemingly concern the selfsame character. But how can they, if, without having any particular object in mind, we each independently merely make believe that 'Anna Karenina' is a genuine name that refers to some real entity?

Writers such as Walton and Currie suggest somewhat different answers to such questions. Common to their answers, however, is the view that the most we can make believe - in propositional content - is the existential general claim that there is a woman who is named 'Anna Karenina', and so on. We cannot literally make believe *of*

a concrete, particular fictional character that *it* is so-and-so because, actually, there are no such characters. This last point can be supported by an argument of Kripke's in *Naming and Necessity* (1980). Our make-believe is not *of* any actual, concrete, particular entity genuinely referred to by 'Anna Karenina'. Nor (to ignore inconsistent fictions) can it be *of* any purely possible thing that exists in all the possible worlds compatible with what is said in the novel and is picked out from each such world by the properties that the novel assigns to Anna. There is no way to decide which of the many possible entities that have all those properties is the unique, particular referent of the name.

Again, writers in the make-believe tradition may understand our non-make-believe claims to indicate or allude to practices of make-believe. Thus, as Walton suggests, my sober, literal assertion that Anna Karenina is from St Petersburg may in effect claim that to say such a thing in the appropriate game of make-believe is fictionally to speak truly. Finally, when both you and I independently talk of Anna Karenina, or when I compare Anna Karenina to Effi Briest, then, Walton urges, we engage in a shared game of make-believe, or I simply combine the games individually appropriate to *Anna Karenina* and *Effi Briest*. (About such literal assertions and comparisons, Currie offers somewhat different treatments, involving what he holds are further, 'metafictive' and 'transfictive' uses of fictional names.)

Some of these suggestions may succeed. But in the end they deny that there can be a single fictional character - a single entity created by the author - that in actual fact two people independently refer to and describe. They also turn out to deny that it is literally and actually true that one and the same character - one single entity - can occur in different works of fiction (or can be the common focus of different games of make-believe; or of different dreams, illusions and beliefs). If we take seriously the idea that there actually are such characters of fiction, then we are forced back into accepting characters as objects of genuine reference and true predication. If so, then unless some Meinongian or similar view is adopted, it seems that we should combine the abstract-existent-object theory with the make-believe approach. Or perhaps (swallowing popular objections to Lewis's modal realism) we should accept a version of Lewis's counterpart theory (see Lewis, D.K.). We might hold, for example, that 'Anna Karenina' refers to various concrete, particular, non-actual individuals that exist in the possible worlds in which Tolstoi's novel is told as known fact. Those would be the individuals that have all the properties the novel attributes to Anna and that (while not strictly identical to one another) are enough alike to function as counterparts (see Modal logic, philosophical issues in §4). It remains to be seen, however, whether these proposals, or any other ideas for countenancing genuine reference to and predication of characters of fiction, can be successfully defended.

See also: Fictional entities; Proper names; Reference

<div align="right">ROBERT HOWELL</div>

References and further reading

Adams, F., Fuller, G. and Stecker, R. (forthcoming) 'The Semantics of Fictional Names', *Pacific Philosophical Quarterly*.(Proposes forgoing fictional characters in favour of talk of fictional names, using the notion of a partial proposition.)

Crittenden, C. (1991) *Unreality*, Ithaca, NY, and London: Cornell University Press. (Defends a Meinongian view against competing theories.)

Currie, G. (1990) *The Nature of Fiction*, Cambridge: Cambridge University Press.(A comprehensive, well-argued make-believe account of the semantics of fiction and related issues. Many valuable references to the earlier literature. Technical in places. Discussed in §§2-3.)

Howell, R. (1979) 'Fictional Objects: How They Are and How They Aren't', *Poetics* 8: 129-77.(Presents difficulties for views rejecting fictional objects. Technical in places.)

Inwagen, P. van (1977) 'Creatures of Fiction', *American Philosophical Quarterly* 14: 299-308.(A forceful defence of an abstract-existent-object view. Discussed in §1.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(A centrally important discussion of reference in general with application to fictional terms such as 'unicorn' and 'Sherlock Holmes'.)

Kroon, F. (1994) 'Make-Believe and Fictional Reference', *Journal of Aesthetics and Art Criticism* 52: 207-14. (Presents problems for make-believe accounts of fictional names; suggests a counterpart theory.)

Lewis, D.K. (1978) 'Truth in Fiction', in *Philosophical Papers*, New York and Oxford: Oxford University Press, 1983, vol. 1, 261-76, with 'Postscripts', 277-80.(A classic pretence account of truth in fiction, with analyses of

'in fiction *f*, so-and-so is the case' claims.)

Parsons, T. (1980) *Nonexistent Objects*, New Haven, CT, and London: Yale University Press.(A careful defence of a Meinongian theory. Technical in parts but with excellent nontechnical discussions. Discussed in §1.)

Routley, R. (1980) *Exploring Meinong's Jungle and Beyond*, Canberra: Research School of Social Sciences, Australian National University.(A broad discussion of Meinongian theories and their applications. Technical in places.)

Russell, B. (1905) 'On Denoting', *Mind* 14: 479-93; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992.(A classic treatment of reference and description.)

Walton, K. (1990) *Mimesis as Make-Believe*, Cambridge, MA: Harvard University Press.(A rich, highly original development of a make-believe theory of all aspects of verbal and nonverbal fictions. Excellent references to the earlier literature. Offers a subtle account of how fictional truths are generated - implicit truths such as 'Anna Karenina has a cerebral cortex' as well as explicit ones. Discussed in §§2-3.)

Wolterstorff, N. (1980) *Works and Worlds of Art*, Oxford: Clarendon Press.(Treats characters as collections of properties. Regards works of fiction as vehicles of special illocutionary acts.)

Woods, J. (1974) *The Logic of Fiction*, The Hague: Mouton.(Treats fictional claims via the sentences occurring in fiction, utilizing substitutional quantification. Technical.)

Holism: mental and semantic

Mental (or semantic) holism is the doctrine that the identity of a belief content (or the meaning of a sentence that expresses it) is determined by its place in the web of beliefs or sentences comprising a whole theory or group of theories. It can be contrasted with two other views: atomism and molecularism. Molecularism characterizes meaning and content in terms of relatively small parts *of the web in a way that allows many different theories to share those parts. For example, the meaning of 'chase' might be said by a molecularist to be 'try to catch'. Atomism characterizes meaning and content in terms of* none *of the web; it says that sentences and beliefs have meaning or content independently of their relations to other sentences or beliefs.*

One major motivation for holism has come from reflections on the natures of confirmation and learning. As Quine observed, claims about the world are confirmed not individually but only in conjunction with theories of which they are a part. And, typically, one cannot come to understand scientific claims without understanding a significant chunk of the theory of which they are a part. For example, in learning the Newtonian concepts of 'force', 'mass', 'kinetic energy' and 'momentum', one does not learn any definitions of these terms in terms that are understood beforehand, for there are no such definitions. Rather, these theoretical terms are all learned together in conjunction with procedures for solving problems.

The major problem with holism is that it threatens to make generalization in psychology virtually impossible. If the content of any state depends on all others, it would be extremely unlikely that any two believers would ever share a state with the same content. Moreover, holism would appear to conflict with our ordinary conception of reasoning. What sentences one accepts influences what one infers. If I accept a sentence and then later reject it, I thereby change the inferential role of that sentence, so the meaning of what I accept would not be the same as the meaning of what I later reject. But then it would be difficult to understand on this view how one could rationally - or even irrationally! - change one's mind. And agreement and translation are also problematic for much the same reason. Holists have responded (1) by proposing that we should think not in terms of 'same/different' meaning but in terms of a gradient of similarity of meaning, (2) by proposing 'two-factor' theories, or (3) by simply accepting the consequence that there is no real difference between changing meanings and changing beliefs.

1 The doctrines

Semantic holism is the view that the meaning of a sentence is determined by its place in the web of sentences comprising a whole theory. Mental holism is the corresponding view for belief content - that the identity of a belief content is determined by its place in the web of beliefs comprising a theory. Sometimes holists advocate a more sweeping view in which the identity of a belief is determined by its relations to a body of theories, or even *the whole of a person's belief system*. In what follows, mental and semantic holism are treated as two aspects of a single view.

Holism can be contrasted with two other views: molecularism and atomism. Molecularism characterizes meaning and content in terms of a relatively small part of the web that many different theories may share. For example, the meaning of 'bachelor' might be said by a molecularist to be 'man who has never married'. And the meaning of 'and' might be given by a molecularist version of inferential role semantics via specifying that the inference from '*p* and *q*' to '*p*', and from '*p*', '*q*' to '*p* and *q*' has a special status (for example, it might be 'primitively compelling', in Peacocke's terms; see Semantics, conceptual role). Atomism characterizes meaning and content in terms of none of the web; it says that sentences and beliefs have meaning or content independently of their relations to any other sentences or beliefs and, therefore, independently of any theories in which they appear.

Note the contrast between the semantic issues that are of concern here and those that concern particular phenomena in particular languages. Semantics in the present sense is concerned with the fundamental nature of meaning and what it is about a person that makes their words mean what they do. We might call the present sense the 'metaphysical' sense. Semantics in the other sense - what we might call the 'linguist's' sense - concerns the issues of how meanings of words fit together to determine the semantic properties and internal structures of sentences. Semantics in the linguist's sense concerns such issues as how many types of pronouns there are and why it is that 'The temperature is rising' and 'The temperature is 60 degrees' does not entail that 60 is rising. There are interactions among the two enterprises, but semantics in the linguist's sense can proceed without taking

much notice of the issue of semantic holism.

2 Motivations for holism

The best-known motivation for semantic/mental holism involves Quine's doctrine of *confirmation* holism, according to which 'Our statements about the external world face the tribunal of sense experience not individually but only as a corporate body' (1953: 41). This view gains its plausibility from the logic of theory revision. An experimental datum confirms (verifies; gives us some reason to believe) a statement only in conjunction with a great number of theoretical ideas, background assumptions about the experiment, and assumptions from logic and mathematics, any one of which could be (and in the history of science often has been) challenged when problems arise (see Confirmation theory). If we combine this confirmation holism with the logical positivist doctrine that the meaning of a sentence is its method of verification or confirmation (see Logical positivism; Meaning and verification), that is, if we combine the doctrine that meaning is confirmation with the claim that confirmation is holistic, we get semantic holism. And this implies that talk of the meaning of a sentence in isolation from other sentences makes no more sense than talk of the meaning of 'of' apart from the contexts in which it occurs.

Positivism and confirmation holism are not the only roads to semantic/mental holism. Another route proceeds from considering how people learn actual scientific theories. For example, one does not learn definitions of 'force', 'mass', 'kinetic energy' or 'momentum' in terms that are understood beforehand, for there are no such definitions. Rather, these terms are learned together (in conjunction with procedures for solving problems). As Quine (1954) and Putnam (1965) argued, local 'definitions' in a scientific theory tend to be mere passing expository devices of no lasting importance for the theory itself. And this is quite ubiquitous in theories, where there is a circle of interdefined theoretical terms none of which is definable in terms outside the theory. This fact motivates Lewis' proposal that scientific terms can be defined functionally in terms of their roles in a whole theory.

3 Must functionalism lead to holism?

Functionalism has become a popular approach in the philosophy of mind generally (see Functionalism). For example, the difference between the belief that one will win the lottery and the desire that one will win the lottery is plausibly a functional difference (a difference in the roles of the states), since one but not the other leads to test-driving a Ferrari. But functionalists go further, claiming that the common *content* of these propositional attitudes can also be functionally defined (in terms of the cognitive roles of states which have these contents in the psychological economy, including links to inputs and outputs). It has often been supposed that the most important feature of the functional role of a belief in determining its content is its role in inference, and for that reason functionalism about content or meaning is sometimes called 'inferential role semantics'. The functional role of a thought includes all sorts of causes and effects that are non-semantic, for example, perhaps depressing thoughts can lower one's immunity, causing one to become ill. Conceptual roles are functional roles minus such non-semantic causes and effects.

A functional theory of the whole mind must make reference to any difference in stimuli or responses that can be mentally significant. The difference between saying 'damn' and 'darn' can be mentally significant (for example, one can have a policy of saying one rather than the other). Your pains lead to 'darn', mine to 'damn', so our pains are functionally different, and likewise our desires to avoid pain, our beliefs that interact with those desires, and so on. So if we functionally define 'pain' in terms of a theory of the whole mind, we are naturally led to the conclusion that two individuals who differ in this way share no mental states. This is why functionalism can lead to holism.

Molecularists object that if you have a fine-grained way of categorizing, you can just coarsen it. But which causes and effects of pain are constitutive and which are not? The form of a solution could be: 'pain = the state constituted by the following causal relations…', where the dots are replaced by a specification of a subset of the mentally significant causal relations into which pain enters. Putnam suggested we look for a normal form for a computational description of pain, and Lycan (1988) and Rey (1997) have suggested that we construct functional theories at different levels, one of which would be suitable to define 'pain' without distinguishing between 'damn' and 'darn'. But after years of discussion, there is no real solution, not even a proposal of something functional common to all and only pains. Lycan and Rey expect the issue to be settled only by an empirical psychology. Moreover, even if one is optimistic about finding a functional definition of pain, one cannot assume that success

will transfer to functionalist accounts of meaning. Success in the case of meaning would seem to require an analytic/synthetic distinction which many have found independently to be problematic.

4 Problems with the analytic/synthetic distinction

Another route to holism arises from considerations involving the analytic/synthetic distinction, that is, the distinction between claims that are true solely in virtue of meaning and claims that depend also on the way the world is. Quineans often hold that the analytic/synthetic distinction is confused. Some philosophers have argued from the idea that there is something wrong with analyticity to holism. We can put the argument in terms of conceptual role semantics. It seems that some inferences (for example, from 'bachelor' to 'unmarried') are part of meaning-constitutive inferential roles, but others (for example, from 'bachelor' to 'dislikes commitment') are not. However, if there is no analytic/synthetic distinction, then there is no principled way to draw a line between inferences that constitute meaning and those that do not (see Analyticity). So, the argument concludes, if some inferences are part of meaning-constitutive inferential rules, *all* inferences are part of meaning-constitutive inferential roles, and this is a form of holism (Fodor and LePore 1992; Devitt 1995).

This argument is of course fallacious. A bald man can have some hairs, and there is no principled way of drawing a line between the number or distribution of hairs on a bald man and on a non-bald man. But one would not conclude that everyone is bald. Failure to find a principled way of drawing a line need not require one or the other extreme.

Still, the argument is onto something. How would the molecularist choose those inferences which are meaning-constitutive if the meaning-constitutive must be analytic rather than synthetic but there is no such distinction? In fact, the problem is more general, and far from being an argument *for* holism, it casts doubt on holism too. If meaning-constitutivity entails analyticity, any view - molecularist or holist - that postulates anything meaning-constitutive is in trouble if there is no such thing as analyticity.

One response to this argument has been to 'challenge' the principle that a statement or inference that is meaning-constitutive is thereby analytic (Block 1993). There are two very different points of view which see a gap between meaning-constitutivity and analyticity.

One approach to finding a gap between meaning-constitutivity and analyticity derives from the views of Quine and Davidson, on which there is no clear difference between a change of meaning and a change of belief (see Radical translation and radical interpretation). The other appeals to narrow contents (see Content: wide and narrow). Narrow contents are contents that are necessarily shared by 'twins', people who are internally as similar as you like, even though their environments differ. Consider the influential example of Putnam's 'twin earth' which is a planet identical to earth in every respect except that wherever earth has H_2O, it has a superficially similar but chemically different substance, XYZ. Arguably, I and my twin on Putnam's twin earth share a narrow content for 'water' despite the different referents of our words. If meaning is narrow in this sense, it is false that meaning-constitutive sentences or inferences are thereby analytic if meaning is narrow. The narrow contents which constitute meaning *themselves* are neither *true* nor *false* and hence cannot be true in virtue of meaning. For example, let us suppose that my twin and I accept the propositions that we express with 'Water contains hydrogen'. My belief has a true wide content, my twin's has a false wide content, but the narrow content has to be the same (since we are twins) and is therefore neither true nor false (see Putnam, H. §3). Further, we can even imagine a twin earth in which a putative meaning-constitutive inference is invalid. If there is any inference that is a good candidate for analytically defining 'water', it is the inference from 'water' to 'liquid'. But consider a twin earth on which 'water' is used as here to refer to H_2O, but where water is very rare, most of the substances referred to as 'liquids' being granular solids that look like liquids. So 'Water is a liquid' as said by on this twin earth is *false*, even though it is true in our mouths. Perhaps it will be said that what is analytic is not 'Water is a liquid' but 'Water has a liquidish look and feel'. But it is easy to imagine circumstances in which the look and feel of water changes. Perhaps what we should be looking for is not a narrow content that is *true* in virtue of meaning but one that is only *assertible* in virtue of meaning. But it is part of our commitment in the use of natural kind terms that the world plays a part in determining truth-values, so we must regard any appearance of warrant solely in virtue of meaning as superficial.

5 The problem of disagreement and translation

Holism has some weird-sounding consequences. Suppose we say that all of a sentence's inferential links (within a theory or body of theories) are included in its set of meaning-constitutive inferential roles. But what sentences I accept influences what I infer, so how can I reason so as to change my own mind? If I accept a sentence, say, 'Bernini stole the lead from the Pantheon', and then later reject it, I thereby change the inferential role of that sentence, so the meaning of the sentence that I accept is not the same as the one that I later reject. So how can I reason about which of my beliefs should be given up? Along similar lines, one can argue that no two people ever agree or disagree, and that we can never translate anything perfectly from one language to another. The holist owes us a way to reconcile such conclusions with common sense. This section will explore three holistic responses.

Harman (1973) and Block (1986) have argued that we can avoid the problem by replacing the dichotomy between agreement and disagreement with a gradient of similarity of meaning, perhaps multi-dimensional. If I first accept and then reject 'Bernini stole the lead from the Pantheon', it is not as if I have rejected something *utterly unrelated* to what I earlier accepted. This position profits from the analogy with the ordinary dichotomy between believing and disbelieving. Reasoning with this dichotomy can lead to trouble, trouble that is avoided if we substitute a graded notion for the dichotomy. For example, I can have a low degree of belief in a long conjunction even though I have a high degree in each of the conjuncts. But if we put this in terms of the dichotomy between believing and disbelieving, we say that I could believe each conjunct while disbelieving the conjunction, and that is a contradiction. The proposal, then, is that we substitute a graded notion of similarity of meaning for the ordinary notion of same/different meaning. It must be conceded, however, that there are no specific suggestions as to what the dimensions of similarity of meaning are or how they relate to one another.

This approach can be combined with the aforementioned 'two-factor theory', according to which meaning consists of an internal holistic factor and a non-holistic purely referential factor (see Semantics, conceptual role §2). For purposes of translation and communication, the purely referential factor plays the main role in individuating contents. For purposes of psychological explanation, the internal factor plays the main role (see Loar 1987).

There is another (compatible) holistic response to the problem of disagreement which is associated with the views of Quine, Davidson and Putnam, namely that there is something wrong with the terms in which the problem is posed. They explicitly reject the very distinction between disagreeing and changing the subject that is presupposed by the statement of the problem. Putnam (1988) and Stich (1983) have argued, along these lines, that translation is not an objective process; it depends on subjective value-laden decisions as to how to weigh considerations of similarity in reference and social and functional role. It is controversial whether this Quinean response avoids the problem of disagreement only by rendering meaning something unsuitable for science.

Another holistic response is exemplified by Lewis' observation (1995) that there is no need to suppose that a satisfier of a functional description must fit it perfectly - fitting *most* of it is good enough. Lewis proposes that in framing the functional roles, we replace the set of inferences that are the basis for a functionalized account of belief with the disjunction of all the conjunctions of most of them. Thus, if we think there are three inferences, *A*, *B* and *C*, that are closely linked to the meaning of 'if', we might define 'if' as the relation that satisfies either $A \& B$ or $A \& C$ or $B \& C$. (Of course, we thereby increase the danger that more than one relation will satisfy our definition.) Then disagreement will be possible between people who accept most of the inferences that define their subject matters.

I have just been canvassing holistic responses to the problem, but of course atomism and molecularism are also responses. Fodor's version of atomism (1987) construes meanings as purely referential. Fodor goes so far as to insist that there could be punctate minds; minds that have only one belief. This view must, however, find some way of accommodating the insights that motivate holism.

6 Psychological laws

Fodor and LePore (1992) object to holistic accounts of mental content on the ground that they would preclude psychological laws, for example, the belief that one is in immediate danger causes release of adrenaline. According to holism, there is no such thing as 'the' belief that one is in immediate danger because the belief that you designate in this way is not quite the same as the belief that I designate in this way. Beliefs are too fine-grained to be referred to in this way (see Belief; Propositional attitudes). One strategy for dealing with this issue is to observe

that many candidate psychological laws can generalize about contents without actually specifying them. Consider this candidate for a law: for *any* action *a* and any goal *g*, if one wants *g* and also believes that *a* is required for *g*, then one will try to do *a*. This is a universally quantified law (because of the role of 'any'), albeit a trivial one. Universally quantified laws are a good scientific bet, and these can involve holistic content. By quantifying over goals, one can state laws without committing oneself to two agents ever having exactly the same goal. The point just made says that the holist can allow one kind of psychological law (the quantified kind) but not another (the kind that mentions specific contents such as the belief that one is in danger). But the holist may go further, arguing that there is something wrong with the putative laws of specific contents. The point is that 'The belief that one is in immediate danger causes release of adrenaline' stands to psychological law as 'Large slippery rocks on mountain-tops can damage cars on roads below' stands to physical law. Laws should quantify over such specific items, not mention them explicitly (see Laws, natural; Explanation).

However, Fodor and LePore are right that any particular type of holistic state will exist only rarely and transiently. In this respect, holistic mental states are like the states of computers (see Mind, computational theories of). A total computer configuration as specified by the contents of every register in the internal memory and every cell on the hard disk will occur only rarely and transiently. There are deterministic laws of the evolution of total computer states, but they deal with such transient states. So psychological explanation will have to be seen by holists as like explanation of what computers do, in part a matter of fine-grained laws of the evolution of systems, in part coarse-grained accounts of how the systems work that do not have the status of laws.

7 Narrow-content holism

There is a great deal of controversy about whether there is such a thing as narrow content or meaning, but if narrow content exists, there is good reason to think it is holistic. We have already seen one reason having to do with the fact that there is no analytic/synthetic distinction for narrow content. But there is another reason as well that focuses on change of narrow content with learning. Putnam (1983) and Block (1994) give an argument that uses some relatively uncontroversial premises about identity and difference in narrow content at a single time to squeeze out a conclusion to the effect that one's narrow contents can be expected to change whenever one receives substantial new information, however trivial. The argument depends on a variant of the famous 'twin earth' example. Consider twins who grow up in different communities where 'grug' is used to denote different substances, beer in one and whisky in the other, but the difference has not made any difference to the twins. At the age of 10, they are as similar as you like, and so the narrow contents of their 'grugs' are the same. By the age of 12, they know as much about 'grug' as teenagers normally know, including the (different) translations of 'grug' into English. One knows that 'grug' in his language is beer, the other that 'grug' is whisky. The argument motivates the claim that their 'grugs' differ in narrow content at 12 despite being the same at 10, so the information that they have acquired (which is designed to be run-of-the-mill) has changed the narrow contents. (But see Devitt 1995 for a reply.)

Issues about holism continue to be at the heart of debate in philosophy of language and mind. In the mid-1960s, it was widely assumed that to be a holist was to be a sceptic about any science of meaning or content, but thirty years later there has been a spirited debate about whether cognitive science can tolerate it.

NED BLOCK

References and further reading

Block, N. (1986) 'Advertisement for a Semantics for Psychology', in P.A. French *et al.* (eds) *Midwest Studies in Philosophy*, vol. 10, *Studies in the Philosophy of Mind*, Minneapolis, MN: University of Minnesota Press. (Discussion of inferential role semantics.)

Block, N. (1993) 'Holism, Hyper-Analyticity and Hyper-Compositionality', *Mind and Language* 8 (1): 1-27. (Defence of holism from a critique by Fodor and LePore.)

Block, N. (1994) 'An Argument for Holism', *Proceedings of the Aristotelian Society*, new series, 94: 151-69. (Concludes that one's narrow contents can be expected to change whenever one receives substantial new information.)

Devitt, M. (1995) *Coming to Our Senses: A Naturalistic Program for Semantic Localism*, Cambridge: Cambridge University Press.(Spirited defence of molecularism.)

Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT

Press.(Defence of an atomistic view of content and meaning.)

Fodor, J.A. and LePore, E. (1992) *Holism: A Shoppers' Guide*, Oxford: Blackwell.(Critique of arguments for holism.)

Harman, G. (1973) *Thought*, Princeton, NJ: Princeton University Press.(Defence of a language of thought. Uses a functionalist point of view.)

Lewis, D. (1995) 'Lewis, David: Reduction of Mind', in S. Guttenplan (ed.) *A Companion to Philosophy of Mind*, Oxford: Blackwell.(Summary and defence of Lewis' views on mind and meaning.)

Loar, B. (1987) 'Social Content and Psychological Content', in R. Grimm and D. Merrill (eds) *Contents of Thought: Proceedings of the 1985 Oberlin Colloquium in Philosophy*, Tucson, AZ: University of Arizona Press.(Argument that narrow content is the only content.)

Lycan, W. (1988) *Judgment and Justification*, Cambridge: Cambridge University Press.(Includes a defence of empirical functional definitions of psychological terms.)

Putnam, H. (1965) 'The Analytic and the Synthetic', in *Philosophical Papers*, vol. 2, *Mind, Language and Reality*, Cambridge: Cambridge University Press, 33-69.(Discussion of the (in)significance of the analytic/synthetic distinction in science.)

Putnam, H. (1983) *Philosophical Papers*, vol. 3, *Realism and Reason*, Cambridge: Cambridge University Press. (A rich development of a conceptual role semantics.)

Putnam, H. (1988) *Representation and Reality*, Cambridge, MA: MIT Press.(Critique of cognitive science views of mind and meaning.)

Quine, W.V. (1953) 'Two Dogmas of Empiricism', in *From a Logical Point of View*, Cambridge, MA: Harvard University Press.(The classic argument against the analytic/synthetic distinction.)

Quine, W.V. (1954) 'Carnap and Logical Truth', in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1976, 70-132.(Rich examination of the role of definition in logic, mathematics and science.)

Rey, G. (1997) *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Oxford: Blackwell. (Defends the prospect of empirical functional definitions of many mental states.)

Stich, S. (1983) *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press.(Argument against appealing to content in psychological explanation.)

Implicature

A term used in philosophy, logic and linguistics (especially pragmatics) to denote the act of meaning or implying something by saying something else. A girl who says 'I have to study' in response to 'Can you go to the movies?' has implicated *(the technical verb for making an implicature) that she cannot go. Implicatures may depend on the conversational context, as in this example, or on conventions, as when a speaker says 'He was clever but poor', thereby implying - thanks to the conventional usage of the word 'but' - that poverty is unexpected given intelligence. Implicature gained importance through the work of H.P. Grice. Grice proposed that conversational implicatures depend on a general principle of rational cooperation stating that people normally try to further the accepted purpose of the conversation by conveying what is true, informative, relevant and perspicuous. The extent and nature of the dependence, and the precise maxims involved, are matters of controversy. Other issues include whether certain implications are implicatures rather than presuppositions or parts of the senses (literal meanings) of the words used.*

1 Speaker meaning and implication

H.P. Grice (1989: chaps 5, 6 and 14) drew an important distinction between what a speaker means or implies, and what a sentence or other expression means or implies. The sentence 'The aeroplane is a mile long' means that the aeroplane is 5,280 feet long, and implies that it is over 3,279 feet long. Nevertheless, it would be most unlikely for a speaker uttering the sentence to mean or imply such things unless he were joking. More likely, the speaker would be exaggerating, meaning simply that the plane is enormous compared with typical planes. A speaker using the sentence in a coded message might mean something completely unrelated to its English meaning.

As a first approximation, word meaning may be characterized as *conventional* speaker meaning. Whereas we could conceivably use 'bachelor' to mean just about anything, we conventionally mean 'unmarried male', which is what the word means. Speaker meaning is determined by the intentions of the individual speaker (see Communication and intention). While the exact intentions required are a matter of controversy, it is plausible that a speaker *S* means that the plane is enormous only if *S* used the sentence to express the belief that the plane is enormous. Typically, *S* would intend to produce that belief in the audience. Speaker implication is *indirect* speaker meaning: meaning one thing (for example, that one cannot go out) by meaning another (that one has to study).

2 Speaker implicature

Grice was the first to systematically study cases in which speaker meaning differs from sentence meaning. He introduced the verb 'implicate' and the cognate noun 'implicature' as technical terms denoting 'the act of meaning or implying something by saying something else' (Grice 1989: chaps 2 and 3). Consider the following dialogue:

(1) Ann: Where can I get petrol?
Bob: There's a petrol station around the corner.

We may suppose Bob to have implied that Ann can get petrol at the petrol station. Nevertheless, Bob did not actually say that Ann can get petrol there. So Bob has 'implicated' it. What Bob said, and therefore did not implicate, is just that there is a petrol station around the corner. By 'saying', Grice meant not mere uttering, but saying *that* something is the case. (As Grice realized, 'say' can be used more or less narrowly. The less speakers are counted as saying, the more they are counted as implicating.)

Example (1) shows that saying is more closely related to conventional meaning than to speaker meaning. The implication that Ann can get petrol at the petrol station is no part of the meaning of 'There's a petrol station around the corner'. If '*e*' means that *p* on a given occasion, then the speaker says that *p*. The converse fails for several reasons. First, a speaker who utters 'Jack and Jill went up the hill' has said that Jack went up the hill; but that is only part of what the sentence means. Second, a speaker who utters 'He is liberal' might have said that President Clinton is liberal; but (being indexical) 'He is liberal' does not mean 'President Clinton is liberal', even on that occasion (see Demonstratives and indexicals). In general, though, an implicature may be described as something the speaker means or implies that is not part of what the sentence literally means.

If '*e*' means that *p*, or if *S* says that *p* by uttering '*e*', then '*p*' is a truth-condition of '*e*', either in general, or as

used on that occasion. Truth-conditions are logical or a priori implications. Assuming particular interpretations, '*p*' is a truth-condition of sentence '*e*' if and only if it is absolutely impossible for '*e*' to be true without '*p*' being true. 'There is a petrol station' is a truth-condition of 'There's a petrol station around the corner'. For the latter cannot be true unless the former is true. But 'Ann can get petrol at the petrol station' is not a truth-condition of 'There's a petrol station around the corner'. The petrol station might be out of petrol or closed. It follows that an implicature need not be a truth-condition of the sentence uttered. The two notions are not mutually exclusive, however, as the following dialogue between an auditor and tax payer illustrates.

(2) Alan: Is it true that you or your spouse is 65 or older or blind?
Bill: I am 67.

By saying 'I am 67', Bill implicated (he implied but did not say) that he or his spouse is 65 or older or blind. Furthermore, he could not be 67 unless he or his spouse were 65 or older or blind. So *S*'s implicature is also a truth-condition of the sentence *S* uttered on that occasion.

Empirically necessary conditions, let us say, are 'natural' implications. 'The power is on' is a natural implication of 'The lights are on'. If '*e*' naturally implies '*p*', then '*e*' means that *p* in Grice's natural sense. Whether natural implications are implicatures depends on the speaker's intentions. In dialogue (3), we would expect Beth to have implicated that the power is on.

(3) Alice: Is the power on?
Beth: The lights are on.

But if Beth does not know what power is, and gives that answer out of desperation, then she would not have implicated that the power is on.

Speakers who know that an implicature is false have misled their audience. But they have not lied unless the implicature happens to be a truth-condition. If Bob knows there is no petrol at the petrol station, then he has misled Ann in dialogue (1). But Bob has not lied, since he did not actually say that there was petrol there. However, if Bill knows that what he has implicated is false in dialogue (2), then he has to be lying.

3 Utterance and sentence implicature

An utterance implicates whatever the utterer implicates. Thus in (1), Bob's utterance implicated that Ann can get petrol. A sentence implicates, roughly, what speakers following linguistic conventions would normally use it to implicate. Despite what Bob implicated in (1), the sentence Bob used does not itself implicate 'Ann can get petrol at the petrol station'. For in most contexts, there would be no impropriety in using 'There's a petrol station around the corner' without that implicature. In contrast, (4a) itself implicates (4b):

(4a) Bill is sick, so he should rest.
(4b) Bill's being sick implies that he should rest.

For a speaker could not properly use (4a) without implicating (4b). And (5a) implicates (5b):

(5a) Some died.
(5b) Not all died.

For speakers would normally use (5a) with that implicature.

4 Conventional versus conversational implicature

Implicatures generated by the conventional meaning of the words uttered, as in (4), are classified as 'conventional' (see, for example, Grice 1989: 25-6). Non-conventional implicatures, as in (5), are termed 'conversational', since they depend on the conversational context. Normally, if one uses a sentence aware of the falsity of some of its conventional implicatures or logical implications, one uses the sentence improperly. If one uses a sentence aware of the falsity of a conversational implicature, one's utterance is at most misleading. Conversational implicatures differ further in being cancellable. The implicature of (5a) can be explicitly cancelled by adding 'indeed, all did'. 'some died, indeed all did' is perfectly consistent, and does not imply 'Not all died' in

any way. In contrast, since 'All died' logically implies 'some died', 'All died, indeed none did' is self-contradictory. 'Bill is sick, so he should rest, but Bill's being sick in no way implies that he should rest' is incoherent in a different way. Conventional implicatures can also be implicitly cancelled by the context. Thus (5a) would not have its customary implication if uttered in full view of a garden in which every single plant was obviously dead. Conversely, conversational implicatures are reinforceable. Conjoining a sentence with one of its conversational implicatures is not redundant, and logically implies what the original sentence implicates; witness 'some died, but not all did'; in contrast, 'All died, and some did' is redundant.

5 Conversational maxims and the cooperative principle

Besides identifying the phenomenon of implicature, Grice formulated a theory in terms of which he classified different sorts of conversational implicature, and tried to explain how they arise and are understood. It is common knowledge, he asserted, that people generally follow certain rules for efficient communication. Grice's rules (1989: 26-9) included one general 'principle of cooperation', and four 'maxims' specifying how to be cooperative:

The cooperative principle. Contribute what is required by the accepted purpose of the conversation.
The maxim of quality. Make your contribution true, so do not convey what you believe false or unjustified.
The maxim of quantity. Be as informative as required, neither more nor less so.
The maxim of relevance. Be relevant.
The maxim of manner. Be perspicuous, so avoid obscurity and ambiguity, and be brief and orderly.

Generalizations of these rules govern rational, cooperative behaviour in general. If I am helping a man change the oil, I will hand him a can of oil rather than a jack (relevance), a barrel of oil (quantity), or fake oil (quality), and I will not take all day doing it (manner). The maxims do not, therefore, have the characteristic arbitrariness of linguistic conventions. Much of the literature is devoted to clarifying and strengthening the maxims, relating them, and extending the list. But Grice's formulations are still dominant.

To sketch how the principles are thought to explain the way implicatures arise and are understood, consider (1). Bob would have infringed the maxim of relevance unless he believed Ann could get petrol at the petrol station. Since Bob was trying to cooperate, his utterance indicated that belief. Thus Bob could imply that Ann can get petrol there and expect Ann to recognize that implication. Speakers using (5a) would violate the maxim of quantity if they knew that all died. So by not making the stronger statement, they can imply that not all died, and expect hearers to recognize the implication. Unlike conventional implicatures, conversational implicatures are cancelled when the maxims are in abeyance. Under cross-examination of an uncooperative witness, (5a) would not implicate (5b).

Some implicatures depend on flouting the maxims, Grice proposed (1989: 30-1). This occurs when what a cooperative speaker says so obviously fails to obey the maxims that the hearer must assume the speaker to be meaning something different. Irony and metaphor appear to depend on flouting the maxim of quality. Hearing 'Fine day, isn't it?' in the middle of a blizzard, we would recognize that the speaker cannot really believe it to be a fine day, and means the opposite. Hearing 'The music danced lightly on his ears', we would recognize that the speaker believes no such thing, and is using metaphor. The speaker who says 'The singer produced a series of sounds closely corresponding to the score of *Oklahoma*', thereby meaning that the singing was terrible, is flouting the maxim of manner.

Many of Grice's theoretical claims about the exact role played by the maxims in implicature are implausible. For example, Grice suggested that *S* conversationally implicates '*p*' only if: (i) *S* is presumed to be observing the cooperative principle ('cooperative presumption'); (ii) the supposition that *S* believes '*p*' is *required* to make *S*'s utterance consistent with the cooperative principle ('determinacy'); and (iii) *S* believes or knows, and expects the hearer *H* to believe, that *S* believes *H* is able to determine that (ii) is true ('mutual knowledge'). The cooperative presumption would fail if, as is possible, Beth implied what she did in case (3) in defiance of Alice's mistrust, or if Beth implicated that she cannot answer the question by saying 'I'm late for an appointment'.

Determinacy fails outright in case (1), as presumably does mutual knowledge: the assumption that Bob believed Ann could at least learn where to get petrol would also reconcile his utterance with the maxims, as would the more fantastic assumption that Bob believed the petrol station was giving away new diesel Volvos. Indeterminacy is of

legendary proportions in figures of speech. Mutual knowledge may even fail independently. In (3), for example, Beth may have contemptuously implied that the power is on, believing that Alice would not be smart enough to figure it out. Some authors have nevertheless strengthened the mutual knowledge condition to say that *S* knows that *H* knows that *S* knows that *H* knows *ad infinitum*, a requirement both psychologically unrealistic and theoretically unnecessary.

While the maxims undeniably provide insight into conversational implicature, their predictive and explanatory power is extremely limited. Consider the account of irony. Grice suggests that *S* implicates 'It is a lousy day' when saying 'Fine day, isn't it?' because otherwise *S* would be violating the maxim of quality and thereby the cooperative principle. But *S* is still violating that principle, by violating the maxim of manner: figures of speech are seldom the most perspicuous ways of communicating information. Consider quantity implicatures. If (5a) implicates (5b) because of the maxim of quantity, then why does (5a) not implicate 'It is not the case that two per cent died'? Why does 'The repairs will take some time' not implicate 'They will not take a long time'? Why does 'Ann said Bill was responsible' not implicate 'Ann did not insist (or predict, or argue, or shout, or…) that Bill was responsible'? Consider finally the sequential interpretation of conjunctions. Grice explained why a conjunction like (6a) implies (6b) in terms of the submaxim 'Be orderly'.

(6a) Bill fell ill and saw the doctor.
(6b) Bill fell ill and therefore saw the doctor.
(6c) Bill saw the doctor and therefore fell ill.
(6d) It is more likely that Bill fell ill than that he saw the doctor.
(6e) Bill did not see the doctor because he fell ill.

But reverse temporal or causal order is just as orderly as forward order (witness 'Bill died; he had cancer'). So why should (6a) not imply (6c)? It would also be quite orderly to express the conjuncts from most to least probable. So why is (6d) not implied? And since '*B* because *A*' is stronger than '*A* and *B*', why is (6e) not implied in virtue of the quantity maxim? (For a comprehensive critique of Gricean theory, see Davis 1998.)

6 Theoretical importance

Difficulties in Gricean theory do not cast doubt on the existence or theoretical importance of implicature. Speakers cannot be fully understood without knowing what they implicate. It does not suffice to know the truth-conditions (or in some cases even the meaning) of all the sentences uttered, or what is said. Implicature must be considered even in truth-conditional semantics. Here are two examples.

Ambiguity. Readily available evidence suggests that the word 'and' is ambiguous in English, having a strong sense (literal meaning) connoting temporal or causal order in addition to the weak sense connoting joint truth alone. But before this can be accepted, the competing hypothesis that the sequential connotation is merely an implicature must be rejected (see Ambiguity §3). Many facts support implicature over ambiguity. For example, 'Bill fell ill and saw a doctor, but not necessarily in that order' does not have a contradictory reading. Furthermore, 'It is not the case that Bill fell ill and saw a doctor' would be interpreted as false when 'Bill fell ill' and 'Bill saw a doctor' are both true, no matter what the sequence of events.

Grice (1989: 47-50) proposed a version of 'Ockham's razor' according to which 'senses are not to be multiplied beyond necessity'. But the same can be said for implicatures. It is often claimed that since implicatures can be accounted for in terms of general psycho-social principles of conversation, the postulation of senses rather than implicatures results in a more complex overall theory. But this overstates the explanatory powers of any known principles of conversation.

Presupposition. '*e*' is said to (semantically) 'presuppose' '*p*' provided the truth of '*p*' is necessary for '*e*' to be either true or false (see Presupposition). Following P.F. Strawson, and opposing Bertrand Russell, many have argued that a sentence like (7a) presupposes (7b):

(7a) Your crimes are inexcusable.
(7b) You have committed crimes.
(7c) Your crimes are not inexcusable.

The presupposition hypothesis is plausible since (7c) seems to imply (7b) just as strongly as (7a) does. Accepting presuppositions seriously complicates logical theory, however. For example, if some declarative sentences are neither true nor false, then standard formulations of the 'law of excluded middle' must be modified. One strategy for avoiding these complications hypothesizes that the implication of the negation is merely a conventional implicature (see, for example, Karttunen and Peters 1979). On this view, (7c) conventionally implicates (7b) without logically implying it. The fact that a sentence may be true when an implicature is false allows the Russellian to account for the felt implications, while insisting that (7c) is true and (7a) false when (7b) is false. Supporting Strawson, though, is the strong intuition that 'Are your crimes excusable or not?' is a loaded question, which cannot be answered if you are innocent. An outstanding problem for either approach is to describe how implications of complex propositions (whether presuppositions or implicatures) are related to those of their components.

Implicature has also been invoked in accounts of lexical gaps, language change, indirect speech acts, textual coherence, discourse analysis and even syntax.

See also: Meaning and communication; Pragmatics; Semantics; Speech acts

WAYNE A. DAVIS

References and further reading

Blakemore, D. (1992) *Understanding Utterances*, Oxford: Blackwell.(A readable introduction to the theory of implicature developed in Sperber and Wilson (1986).)

Davis, W.A. (1998) *Implicature: The Failure of Gricean Theory*, Cambridge: Cambridge University Press.(A comprehensive critique of Grice's theory of implicature, and of similar theories. Argues that conversational implicatures arise not from conversational principles, but from intentions and conventions.)

Gazdar, G. (1979) *Pragmatics: Implicature, Presupposition, and Logical Form*, New York: Academic Press. (Attempts a partial formalization of conversational implicatures, focusing on quantity implicatures, and uses it to provide a pragmatic theory of presupposition.)

Grice, H.P. (1989) *Studies in the Ways of Words*, Cambridge, MA: Harvard University Press.(Contains all of Grice's work on meaning and implicature, plus an introduction and retrospective epilogue.)

Harnish, R.M. (1976) 'Logical Form and Implicature', repr. in S. Davis (ed.) *Pragmatics: A Reader*, Oxford: Oxford University Press, 1991, 316-64.(Presents many problems for Gricean theory, suggesting solutions and extensions.)

Horn, L.R. (1989) *A Natural History of Negation*, Chicago, IL: University of Chicago Press.(A remarkably comprehensive and detailed account of negation, focusing *inter alia* on its effect on implicature, and explanation of a wide variety of linguistic facts about negation in terms of implicature.)

Karttunen, L. and Peters, S. (1979) 'Conversational Implicature', in C.-K. Oh and D.A. Dinneen, *Syntax and Semantics*, vol. 11, *Presupposition*, New York: Academic Press, 1979, 1-56.(A detailed, compelling argument that different sorts of presupposition can be accounted for as different types of implicature, plus a representation of conventional implicature using model-theoretic semantics, specifically, Montague grammar.)

Leech, G. (1983) *Principles of Pragmatics*, London: Longman.(A leading introduction to pragmatics stressing implicature. Develops Grice's suggestion that another maxim is 'Be polite', showing that it motivates many implicatures.)

Levinson, S.C. (1983) *Pragmatics*, Cambridge: Cambridge University Press.(A leading introduction to pragmatics, with an extensive discussion of conversational implicature and presupposition from a Gricean point of view.)

Martinich, A.P. (1984) 'A Theory for Metaphor', *Journal of Literary Semantics* 13: 35-56; repr. in S. Davis (ed.) *Pragmatics: A Reader*, Oxford: Oxford University Press, 507-18.(An extended Gricean treatment of metaphor.)

Sadock, J.M. (1978) 'On Testing for Conversational Implicature', in S. Davis (ed.) *Pragmatics: A Reader*, Oxford: Oxford University Press.(Points out problems with all of Grice's tests for conversational implicature.)

Sperber, D. and Wilson, D. (1986) *Relevance: Communication and Cognition*, Oxford: Blackwell.(Develops a theory of implicature in terms of a single 'principle of relevance' quite distinct from Grice's 'maxim of relevance'.)

Indicative conditionals

Examples of indicative conditionals are 'If it rained, then the match was cancelled' and 'If Alex plays, Carlton will win'. The contrast is with subjunctive or counterfactual conditionals, such as 'If it had rained, then the match would have been cancelled', and categoricals, such as 'It will rain'.

Despite the ease with which we use and understand indicative conditionals, the correct account of them has proved to be very difficult. Some say that 'If it rained, the match was cancelled' is equivalent to 'Either it did not rain, or the match was cancelled'. Some say that the sentence asserts that the result of 'adding' the supposition that it rained to the actual situation is to give a situation in which the match was cancelled. Some say that to assert that if it rained then the match was cancelled is to make a commitment to inferring that the match was cancelled should one learn that it rained. This last view is often combined with the view that indicative conditionals are not, strictly speaking, true or false; rather, they are more or less assertible or acceptable.

1 Indicative and material conditionals

In general an 'indicative conditional' has the form 'If *A* then *C*', where *A* is called the antecedent and *C* the consequent. A central issue is the relationship between the truth-value of a conditional and the truth-values of its antecedent and consequent. This much is immediately plausible: if *A* is true and *C* is false, then the conditional is false. If I say 'If it rained, the match was cancelled', and what happened was that it rained but the match went ahead, then what I say is clearly false. There are three other possibilities: *A* and *C* are both true, *A* is false and *C* is true, and *A* and *C* are both false. There are a number of arguments designed to show that in each of these cases the conditional is true. Here is one. 'If *A* then *C*' is intuitively equivalent to the disjunction 'Either not-*A*, or *A* and *C*'. (Instead of saying that if it rained the match was cancelled, I could have said 'Either it did not rain, or it did and the match was cancelled'.) But the latter is true in all three cases: in each, either the first disjunct ('not-*A*') or the second disjunct ('*A* and *C*') is true. A conditional that is false when its antecedent is true and its consequent false, and true in all other cases, is called a 'material conditional', and is symbolized ' $A \supset C$' (read '*A* hook *C*'). ' $A \supset C$' is definitionally equivalent to 'Not-*A* or *C*'. Hence the argument's conclusion is that indicative conditionals are equivalent to material conditionals. This conclusion has the virtue of validating the two most obviously valid inferences governing conditionals - *modus ponens*: '*A*, if *A* then *C*, therefore, *C*', and *modus tollens*: 'Not-*C*, if *A* then *C*, therefore, not-*A*'.

Despite the appeal of the argument, there are serious problems for the equivalence thesis that indicative conditionals are material conditionals. It entails that any conditional with a false antecedent is true regardless of its consequent, that is, the account validates 'Not-*A*, therefore, if *A* then *C*'. This is implausible. 'If I live in London then I live in Scotland' strikes us as false (it is rather 'If I live in London then I do not live in Scotland' which is true), but because I do not live in London, it is true on the equivalence thesis. Also, it entails that any conditional with a true consequent is true: the account validates '*C*, therefore, if *A* then *C*'. But 'If I live in London, then I live in Australia' strikes us as false even after we learn that I do in fact live in Australia. These two results are known as the paradoxes of material implication ('material implication' being the name of the relation between *A* and *C* when ' $A \supset C$' is true).

2 Possible worlds treatments

An obvious response to the paradoxes is to insist that more than the truth of ' $A \supset C$' is required in order for 'If *A* then *C*' to be true; that the truth of ' $A \supset C$' is a necessary but insufficient condition. One way to do this is to require a connection between antecedent and consequent; to hold that part of what makes 'If it rains, the grass will grow' true is the connection between rain and grass growing. This blocks the paradoxes. However, we sometimes use conditionals precisely to deny that there is a connection between antecedent and consequent. A doctor who says that if you go to bed and take an aspirin you will get better in a week, whereas if you go to work it will take seven days, is saying that there is no connection between going to bed and getting better.

A more promising approach is to require that the material conditional not only be true the way things actually are, but be true in the closest possible worlds where the antecedent is true (see Semantics, possible worlds). Roughly, the account is: 'If *A* then *C*' is true iff ' $A \supset C$' is true, that is, true the ways things actually are *and also* true in

the closest possible worlds where A is true. Another way of saying the same thing is: 'If A then C' is true iff the closest worlds in which A is true are worlds in which C is also true. (This is because any world where A and '$A \supset C$' are true is a world where C is true.) The appeal of this approach derives from the appeal of the idea that when we evaluate a conditional we look at the way things actually are and 'in imagination' add the antecedent and then see whether, with only the changes forced by adding the antecedent, the consequent comes out true.

This approach is very attractive for subjunctive or counterfactual conditionals (see Counterfactual conditionals) but faces a problem when applied to indicative conditionals. (Interestingly, one architect of the possible worlds approach to conditionals, David Lewis (1973), only ever intended it to apply to subjunctive conditionals. Robert Stalnaker, the other architect (1968), intended it to apply to both.) It construes indicative conditionals as being in part about the way things might have been but are not in fact. That seems exactly right for subjunctives. When I say that had I invested in Western Mining ten years ago, I would now be rich, I am saying how things are in some possible but regrettably non-actual world. Indeed, I might express this by saying that had I invested in Western Mining ten years ago, things would now be different - and better - than they actually are. But I cannot say this in the indicative mood: 'If I invested in Western Mining ten years ago, things are different from the way they actually are' is nonsense. Or consider the difference between 'If Oswald had not shot Kennedy, Kennedy would have won a second term' and 'If Oswald did not shoot Kennedy, Kennedy won a second term'. The reason we reject the second, despite accepting the first, is that we know that the way things actually are Kennedy did not live to win a second term. Subjunctive conditionals typically concern non-actual worlds, whereas indicative conditionals concern the actual world under various hypotheses about what it is like. Thus we say that if Oswald did not shoot Kennedy, someone else did, because that is the only thing to think about the actual world (in which we know that Kennedy was shot) given that Oswald was not responsible.

3 Adams and Lewis

What about the acceptance or justified assertion conditions, as opposed to truth-conditions, for indicative conditionals? Many, including particularly Ernest Adams (1975), have urged that it is justified to assert 'If A then C' to the extent that C is probable given A. Because the probability of C given A - in symbols, $Pr(C|A)$ - is (roughly) the probability we ascribe to C on learning A (see Probability, interpretations of §5), this suggestion fits nicely with the plausible idea that one is prepared to assert 'If A then C' to the extent that one is prepared to infer C on learning A. It also explains our reluctance to assert together 'If A then C' and 'If A then not-C' when A is consistent, for when A is consistent, $Pr(C|A) = 1 - Pr(\text{not-}C|A)$, so they cannot be high together.

Can we explain this assertibility condition in terms of truth-conditions? David Lewis (1976) showed that it cannot be the case that the truth-conditions of 'If A then C' are such that $Pr(\text{If } A \text{ then } C) = Pr(C|A)$. For then this equality would hold for all probability functions Pr, and that leads to trouble as follows:

$$Pr(\text{If } A \text{ then } C) = Pr(\text{If } A \text{ then } C|C).Pr(C) + Pr(\text{If } A \text{ then } C|\text{not-}C).Pr(\text{not-}C),$$

by expansion by cases. But if $Pr(\text{If } A \text{ then } C) = Pr(C|A)$ holds for all Pr, then it holds for $Pr(-|C)$ and $Pr(-|\text{not-}C)$, as the class of probability functions is closed under conditionalization. This means that

$$Pr(\text{If } A \text{ then } C|C) = Pr(C|A.C)$$
$$Pr(\text{If } A \text{ then } C|\text{not-}C) = Pr(C|A.\text{not-}C)$$

(where $Pr(C|A.C)$ is the probability of C given the conjunction of A and C). Hence we have

$$Pr(\text{If } A \text{ then } C) = Pr(C|A.C).Pr(C) + Pr(C|A.\text{not-}C).Pr(\text{not-}C)$$
$$= 1.Pr(C) + 0.Pr(\text{not-}C)$$
$$= Pr(C)$$

But then, by the claim under discussion, $Pr(C) = Pr(C|A)$. This is a 'reductio', for in general the probability of C is not independent of that of A. (This is the essence of the simplest of the proofs Lewis offered. For the substantial developments see Lewis (1986).)

Many respond to this proof by holding that indicative conditionals do not have truth-conditions at all; they only have assertion or acceptance conditions. This theory undoubtedly has its attractions but faces two problems. What

should be said about the powerful intuition that a conditional with a true antecedent and a false consequent is false? And how should the relevant notion of assertibility be elucidated? Not in terms of likelihood of truth, obviously.

An alternative strategy, called the supplemented equivalence theory, is to return to the equivalence thesis and argue that there is a convention governing the assertion of 'If A then C' to the effect that it should only be asserted when it would be right to infer C on learning A. This convention is like that governing the use of 'but'. 'A but C' has the same truth-conditions as 'A and C', but the use of the former conventionally implicates, in H.P. Grice's terminology, a contrast (see Grice 1975; Implicature). Likewise, runs the suggestion, 'If A then C' has the same truth-conditions as '$A \supset C$' but its use carries the implicature that my reasons for '$A \supset C$' are such that it would be right, on learning A, to infer C (that is, to use *modus ponens*). Now that will be the case just if the probability of '$A \supset C$' would not be unduly diminished on learning that A is true - otherwise it would not then be available as a probably true premise to combine with A on the way to inferring C. It follows that it will be right to assert 'If A then C' to the extent that (1) '$A \supset C$' is probable, and (2) '$A \supset C$' is probable given A. It is an elementary exercise in probability theory to show that this two-fold condition is satisfied to the extent that $Pr(C|A)$ is high. The supplemented equivalence theory therefore explains the assertibility condition noted by Adams.

What do supplemented equivalence theorists say about our earlier example 'If I live in London, then I live in Scotland', which comes out true on the equivalence theory and so on the supplemented equivalence theory? They say that it seems false not because it is false but because it has very low assertibility - *Pr*(I live in Scotland/I live in London) = 0 - and in general that our intuitions about truth and validity of inferences involving indicative conditionals are governed by responses to assertibility rather than truth.

Warning: almost everything about indicative conditionals is controversial, including whether they are well labelled by the term 'indicative', and some even deny the validity of *modus ponens*!

See also: Logical and mathematical terms, glossary of; Relevance logic and entailment

<div align="right">FRANK JACKSON</div>

References and further reading

Adams, E. (1975) *The Logic of Conditionals*, Dordrecht: Reidel.(A detailed defence of an approach to conditionals that denies them truth-conditions and shows how to develop a logic of conditionals based on his views about their assertibility.)

Appiah, A. (1985) *Assertion and Conditionals*, Cambridge: Cambridge University Press.(A detailed exposition of an account of indicative conditionals in terms of assertibility, set within the philosophy of mind.)

Edgington, D. (1995) 'On Conditionals', *Mind* 104: 235-329.(Excellent survey article that contains interesting material in its own right; includes a comprehensive and up-to-date reading list.)

Grice, H.P. (1975) 'Logic and Conversation', in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991.(Explains implicature and the distinction between conversational and conventional implicature.)

Jackson, F. (1987) *Conditionals*, Oxford: Blackwell.(A detailed defence of the supplemented equivalence theory with criticisms of various alternative accounts of indicative conditionals.)

Lewis, D.K. (1973) *Counterfactuals*, Oxford: Blackwell.(The book which, along with Stalnaker's paper, launched the possible worlds approach to conditionals.)

Lewis, D.K. (1976) 'Probabilities of Conditionals and Conditional Probabilities', in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991.(Gives the full version of the proof sketched in this entry.)

Lewis, D.K. (1986) 'Probabilities of Conditionals and Conditional Probabilities II', in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991.(More on the proof sketched in this entry.)

Stalnaker, R. (1968) 'A Theory of Conditionals', in F. Jackson (ed.) *Conditionals*, Oxford: Oxford University Press, 1991.(The paper which, along with Lewis's book, launched the possible worlds approach to conditionals.)

Indirect discourse

Indirect discourse is a mode of speech-reporting whereby a speaker conveys the content of someone's utterance without quoting the actual words. Thus, if Pierre says, 'Paris est belle', an English speaker might truly say,

(1) Pierre said that Paris is beautiful.

In English, sentences of indirect discourse often have the form 'A said that s', where 'A' refers to a person and 's' is often called the 'content sentence' of the report.

Sentences of indirect discourse have been classed with attributions of belief (and other psychological states) in view of an apparent conflict with the 'principle of the intersubstitutability of coreferring terms', which states that the truth-value of a sentence does not alter if one term in a sentence is replaced with another referring to the same thing. If (1) is true and 'Paris' and 'the City of Light' refer to the same thing, (2) may still be false:

(2) Pierre said that the City of Light is beautiful.

1 Frege's theory

In 'indirect discourse' a speaker reports someone's utterance without quoting the actual words used (often using the form '*A* said that *s*'). Sentences of indirect discourse give rise to an apparent conflict with the 'principle of the intersubstitutability of coreferring terms', which states that the truth-value of a sentence does not alter if one term is replaced with another which refers to the same thing. Gottlob Frege (1892) put forward an important theory to deal with this problem. He proposed that 'said' alters what the words in the 'content sentence' ('*s*' above) refer to. For example, consider the following statements:

(1) Pierre said that Paris is beautiful.
(2) Pierre said that the City of Light is beautiful.

'Paris' and 'the City of Light' both usually refer to the actual city, Paris. The phrase 'is beautiful' usually refers to the set of beautiful things. But when words appear after the phrase 'said that', they refer to the meanings they normally express - which Frege called their 'senses' - which are not the same as their referents (see Frege, G. §3).

The sense of an expression is a 'mode of presentation' of its referent. The sense associated with 'the City of Light' presents Paris in a particular way: as the City of Light. To understand an expression is to grasp its sense (see Sense and reference). Frege called the sense of a sentence a 'thought'. To have a propositional attitude is to stand in a certain relation to a thought. For example, to believe that Paris is beautiful is to stand in the believing relation to the thought that Paris is beautiful. Frege claimed that senses and thoughts are not psychological, mind-dependent things, but rather are Platonic, abstract objects. If two people believe that Paris is beautiful, then it is the very same thought that they both have. And this thought would exist even if there had been no minds (see Frege, G. §4).

The Fregean theory explains away the apparent failures of intersubstitution: the 'senses' of 'Paris' and 'the City of Light' are different. Hence the words refer to different things in (1) and (2) and so no violation of the substitution principle occurs.

2 Davidson's theory

Donald Davidson (1968/69) has provided a very different account of indirect discourse. Davidson proposed that, at the level of logical form (see Logical form), the analysis of speech reports should be composed of two separate sentences, the first ending in the demonstrative pronoun 'that'. The analysis of (1) above would then be:

(3) Pierre said that. Paris is beautiful.

On Davidson's theory the two sentences are logically distinct (not connected by any logical particle, such as 'and'). However, on specific occasions of use the demonstrative 'that' ending the first sentence refers to the reporter's utterance of the second.

According to this theory, the direct object of the verb 'said' is the reporter's own token utterance. So, suppose that

someone, *A*, utters (1). This speech act is composed of two token utterances: one of 'Pierre said that' and one of 'Paris is beautiful'. Call this latter utterance '*U*'. Then, what Pierre would be reported as saying is *U*.

This feature is surprising, since it does not seem that Pierre could have said *U* (the reporter's utterance) itself. But Davidson provides an account of what 'said' means which addresses the difficulty. He introduces the notion of 'samesaying': roughly, two individuals are samesayers if and only if they produce utterances which have relevantly similar semantic contents. Davidson has provided a general theory of meaning and interpretation which underwrites his appeal to semantic contents in this context (1984). We are then to understand (1) along the following lines:

(4) There was some utterance, *u*, of Pierre's, and *u* and my next utterance make us samesayers. Paris is beautiful.

The idea is that the relation expressed by 'said' in *A*'s utterance of (1) can be understood in terms of the more complex relation spelled out in (4). But (4) is not intended to be any part of a technical semantic theory. It is just an informal, extra-theoretic guide to readers to help them understand the theory. In a semantic theory in Davidson's style (see Davidson, D. §4), the axiom for 'said' might be:

(5) $(x)(y)$('said' is true of $\langle x, y \rangle$ iff x said y).

The point is that since 'said' is supposed to relate reportees to utterances of reporters, some informal explanation of this saying relation is required. (4) provides this explanation.

Davidson's theory explains the apparent failures of substitution such as that illustrated by (1) and (2). In an utterance of (1) the main verb is 'said' and its syntactic object is the demonstrative 'that'. The ensuing utterance of 'Paris is beautiful' only serves as a referent for the demonstrative. It is not in any other way a semantically relevant component of the report. An utterance of (1) and an utterance of (2) resemble each other in that each has the form 'Pierre said that'. But the referents of the two instances of 'that' are different utterances. So (1) does not entail (2). No violation of the intersubstitutability of coreferring terms occurs. The latter principle still applies to the content sentences of (1) and (2):

(6) Paris is beautiful.
(7) The City of Light is beautiful.

(6) does entail (7), given the coreference of 'Paris' and 'the City of Light'. But this has no direct bearing on the logical relations between (1) and (2).

Davidson's theory differs importantly from Frege's. The former does not entail any shifts of reference for the words in the content sentence, which has exactly the same semantic features as it would have if uttered in isolation. This feature of the theory implements a natural and elegant idea of how speech reporting works. When *A* reports on the speech of *B*, *A* produces a new sentence, *S*, as content sentence of the report. *S* in *A*'s mouth expresses the content of *B*'s speech act. The utterance of *S* then serves as a model of *B*'s original speech act. And this exhausts its role.

Various objections have been made to Davidson's proposal. Only two are discussed here. The first objection is due to Tyler Burge (1986). Burge points out that on Davidson's theory, an utterance of (8) below would be true only if there were some utterance of the reporter's and some utterance of Pierre's that made them samesayers:

(8) Pierre said something.

This result follows from the functioning of the word 'said' on the theory. But the result is wrong: I might utter (8) truly without myself ever making any utterance that would make Pierre and me samesayers. The objection is perhaps not conclusive. But Burge points out that a solution to the difficulty would be to allow the objects of speech reports to be abstract objects, rather than token utterances, and offers several further arguments that this solution is correct. It seems natural to take the relevant abstract objects to be sentence-types. If this suggestion is adopted, then Burge's objection forces no major revisions of Davidson's theory. The original account of 'said' in terms of samesaying no longer applies. But an analogous explanation of the functioning of 'said' remains available. Instead of understanding (1) along the lines of (4), we would understand it along the lines of:

(9) There was some sentence, *s*, which Pierre uttered, and *s* meant in Pierre's mouth then what the following means now in mine: Paris is beautiful.

The second objection is due to James Higginbotham (1986). On Davidson's account, (10) and (11) ought to have identical semantic features:

(10) Every girl said that her mother was kind.
(11) Every girl said that. Her mother was kind.

This is because (11) merely makes explicit the analysis that Davidson would attribute to (10). However, (10) and (11) are semantically distinguishable. (10) is ambiguous: on the first reading, (10) would be true if and only if each girl said of some specific contextually indicated person, *x*, that *x*'s mother was kind; on the second reading, it would be true if and only if each girl said that her own mother was kind. But (11) is only subject to the first reading. This objection shows that Davidson's theory is not correct as it stands, and will have to modified in very serious ways if it is to accommodate interactions between words inside and outside content sentences.

See also: Logical and mathematical terms, glossary of; Propositional attitude statements

GABRIEL SEGAL

References and further reading

Burge, T. (1986) 'On Davidson's "On Saying That"', in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, New York: Blackwell, 191-208.(A critical article on Davidson's theory of indirect discourse.)

Davidson, D. (1968/69) 'On Saying That', *Synthese* 19: 130-46.(The classic work of Davidson's theory of indirect discourse, including criticisms of Frege's theory.)

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.(A collection of Davidson's articles in the philosophy of language, including some seminal works on the theory of meaning and interpretation.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 1962, 56-78; 3rd edn, 1980.(The first clear statement of the problems of apparent failures of intersubstitutability from the standpoint of technical semantic theory. One of the most influential articles in philosophy of language.)

Geach, P.T. (1957) *Mental Acts*, London: Routledge & Kegan Paul.(A general work in the philosophies of mind and language, including extended discussion of the relationship between saying and propositional attitudes such as believing or judging.)

Higginbotham, J. (1986) 'Linguistic Theory and Davidson's Program in Semantics', in E. Lepore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, New York: Blackwell, 29-48.(Argues for the incorporation of Davidson's main technical innovations in philosophical semantic theory into empirical linguistics and includes a sophisticated theory of indirect speech that retains some core aspects of Davidson's proposal but avoids the main objections to it.)

Scheffler, I. (1954) 'An Inscriptional Approach to Indirect Quotation', *Analysis* 10: 83-90. (An influential nominalist account of indirect discourse.)

Segal, G. and Speas, M. (1986) 'On Saying ðČbepsi;t', *Mind and Language* 1 (2): 124-32. (An article criticizing Davidson's theory of indirect discourse on empirical grounds.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press, 1964.(A general work in systematic philosophy including arguments critical of semantic theories, such as Frege's, that use a mentalistic or abstract notion of meaning; and discussion of various approaches to indirect discourse and propositional attitude statements.)

Intensionality

The truth or falsity of many sentences depends only on which things are being talked about. Within intensional contexts, however, truth values also depend on how *those things are talked about, not just on which things they are. Philosophers and logicians have offered different analyses of intensional contexts and the behaviour of terms occurring within them.*

The extension of a term is the thing or things it picks out: for instance, the extension of 'the Big Dipper' is the stellar constellation itself. The intension of a term can be thought of as the way in which it picks out its extension. 'The Big Dipper' and 'the Plough' have the same extension - a particular constellation - but pick it out in different ways. The two terms have different intensions. There is much debate about how we should actually understand the notion of an intension (see Intensional entities).

For many classes of sentence, the substitution of one term for another with the same extension leaves the truth value unchanged, irrespective of whether the two terms also have the same intension. This is called intersubstitution *salva veritate* (preserving truth). For instance, if it is true that the Big Dipper consists of seven stars, then it is also true that the Plough consists of seven stars. In a number of important cases, however, terms with the same extension but different intensions are not intersubstituable *salva veritate*. Consider, for instance:

1 Tom believes that the Big Dipper consists of seven stars.
2 Tom believes that the Plough consists of seven stars.

It is quite possible that Tom mistakenly believes that 'the Plough' refers to a totally different constellation consisting of only six stars, in which case (1) could be true and (2) false. Co-extensive terms cannot be intersubstituted *salva veritate* within the scope of the verb 'believes'. Such contexts are called intensional.

Other verbs which, like 'believe', refer to propositional attitudes - verbs like 'hope', 'desire', 'fear' - also create intensional contexts (see Propositional attitude statements). Propositional attitudes are *intentional* states (note the 't') which has led some to consider the possible connections between inten*S*ionality and inten*T*ionality (see Intentionality §2). The two must be distinguished carefully, however, not least because intensional contexts are also created by non-psychological terms, most importantly modal terms like 'necessarily' and 'possibly'. The intensionality of modal contexts leads to particular difficulties when they also involve quantification (see Modal logic, philosophical issues in §3); these problems have led to controversies over the interpretation not only of modal terms but also of the quantifiers (see Modal operators; Quantifiers, substitutional and objectual §1).

The failure of intersubstitutability *salva veritate* raises questions about the behaviour of terms in intensional contexts. Frege (1892) argues that, in intensional contexts, the term 'the Big Dipper' in (1) refers not to the Big Dipper as usual but to its own intension (or, in Frege's terminology, its sense; see Indirect discourse §1; Proper names §5; Sense and reference §5). Quine (1956, 1961) describes intensional contexts as *referentially opaque*, arguing that terms occurring within them do not refer at all (see Propositional attitude statements §2; Modal operators §1). Davidson's (1969) *paratactic* analysis of propositional attitude statements and of indirect discourse, another intensional context, tries to preserve our intuition that terms in intensional contexts work in the same way as in any other contexts (see Indirect discourse §2; Propositional attitude statements §3).

Many philosophers think that logic should have nothing to do with intensions. But the intensionality of many natural language contexts is hard to deny, and the development of intensional logics suggests that a more tolerant attitude might well pay off (see Intensional logics).

See also: Concepts; Reference; Semantics, possible worlds

SIMON CHRISTMAS

References and further reading

Davidson, D. (1969) 'On Saying That', *Synthèse* 19: 130-46; repr. in D. Davidson and J. Hintikka (eds) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 158-74.(Davidson's paratactic analysis of indirect discourse.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black as 'On Sense and Reference', *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford: Blackwell, 1962, 56-78.(Frege's seminal discussion of failures of intersubstitution and distinction between sense and reference.)

Quine, W.V. (1956) 'Quantifiers and the Propositional Attitudes', *The Ways of Paradox*, Cambridge MA: Harvard University Press, 1976.(Quine on referential opacity.)

Quine, W.V. (1961) 'Reference and Modality', *From a Logical Point of View*, Cambridge, MA: Harvard University Press, 2nd edn, 1980; repr. in L. Linsky (ed.) *Reference and Modality*, Oxford: Oxford University Press, 1971.(Quine on the referential opacity of modal contexts. The Linsky collection contains some fairly difficult discussions.)

Searle, J.R. (1983) *Intentionality*, Cambridge: Cambridge University Press.(Chapter 1 contains a useful discussion of the relationship between inten*S*ionality and inten*T*ionality.)

Interpretation, Indian theories of

Need for interpretation of texts was felt already during the ancient period of Vedic texts in India. Vedic texts were orally transmitted for over a thousand years. During this period, the change in locations of people reciting the texts and the mother tongues of the reciters led to a widening gap between the preserved sacred texts and their interpreters. Additionally, there was a notion that the sacred language was a mystery which was only partially understood by the common people. This led to the early development of exegetical tools to assist the interpretation of the sacred literature.

Later, grammarians and etymologists developed sophisticated exegetical tools and theories of interpretation. These generally led to a deeper understanding of the structure of language. The priestly tradition developed its own canons of interpretation, which are manifest in the system of Mīmāṃsā. Here we have the first fully developed theory of discourse and context.

The categories developed by Mīmāṃsāwere used by other schools, especially by the school of Dharmaśāstra, or Hindu religious law. Both Mīmāṃsāand Dharmaśāstra created sets of hierarchical principles for authoritative guidance in interpretation. Other philosophical and religious traditions developed categories of their own to deal with problems of interpretation. A major problem was created when the literature accepted as authoritative by a tradition contained apparently contradictory passages. The traditions had to deal with this problem and find ways of explaining away those passages which did not quite fit with their own view of truth. For this purpose, a whole set of categories were employed. At a later period, several ingenious principles of interpretation were used for texts in general. Here, significant contributions were made by the traditions of Sanskrit grammar and poetics.

1 Text and interpreter

Concern for the interpretation of the scriptural language was already being voiced in India from the earliest period of the Vedic scriptures (*c.*1500 BC). The primary reason for such concern lies in the real or perceived distance between the language of a sacred text and the contemporary language of the person trying to comprehend and use that text. Real distance between the language of a sacred text and the language of its interpreter or user can be caused by various factors, such as temporal and geographical distance, linguistic change in the sacred language, shifts in the mother tongues of the text's preservers, and corruption and loss in the text in the course of oral or written transmission. The middle Vedic texts already express a concern that the users of language are unable to pronounce proper accents. The loss of this ability contributed to improper use of sacred formulas, as well as to misinterpretation of their meaning. We are told a story about a demon who seeks a son who will kill the god Indra. He should ask for a son who will be *indraśatrú*, 'killer of Indra'. However, the ignorant demon uses a wrong accent and asks for a son who will be *índraśatru*, 'he whose killer is Indra'. This story is later repeated by Sanskrit grammarians and etymologists to illustrate the importance of a proper understanding of accents. Similarly, in the oral tradition, long stretches were memorized and recited in an unbroken fashion. Besides actual cases of mutilation and alteration of the text, such recitation resulted in multiple possibilities for dividing up the text.

Ideas regarding the nature of the sacred language contributed an additional feeling of distance between text and interpreter. The *Ṛg Veda* tells us that language, in reality, is measured in four quarters, all of which are known to the wise, insightful priests. However, three quarters of speech are hidden away from common people, who have access only to a single quarter. The goddess of speech is said to reveal her true form only to a person of her own choice. Another important notion is that the gods do not like the direct mode of expression (*pratyakṣa*). They would rather have sacred expressions in an indirect mode (*parokṣa*). Middle Vedic prose texts often explain the so-called indirect expressions by paraphrasing them with expressions which have structures more transparent to their audience.

2 Early tools to assist interpretation

In the middle (*c.*1000 BC) and late (*c.*700-500 BC) Vedic periods, we find a growing concern for proper understanding of older texts and the beginning of concerted efforts to develop technical tools towards that goal. For the Brāhmaṇa texts, the highest perfection (*rūpasamṛddhi*) of ritual results when the recited Vedic text matches the ritual action. This places a high premium on comprehending the meaning of the recited texts before

one decides how and when they are to be used in ritual. The texts repeatedly state that the man who performs a ritual 'knowingly' (*ya evaṃ vidvān*) will be rewarded.

There was also a serious effort by priest-scholars such as Māṇḍūkeya and Śākalya to put together the Padapāṭhas, 'word-by-word' versions of Vedic texts. The late Vedic text *Aitareya-Āraṇyaka* speaks of *nirbhuja* ('unsegmented') versions versus *pratṛṇṇa* ('segmented', 'broken-down') versions. Later on, the segmented versions were used as the basis for reproducing continuous versions (*saṃhitā*) which were deemed to match the originals. These activities provided an enormous opportunity to build formal tools to assist interpretation, a task taken up later by the grammarians, etymologists and phoneticians of ancient India.

3 Etymologists and grammarians

The goal of the etymologists was to provide insight into the meaning of an expression (*nirvacana*) by providing an analytical breakdown of it. Yāska, the author of *Nirukta (Analytical Exegesis*, *c.*500 BC), the only surviving ancient text on etymology, says that comprehension of meaning is the highest goal of his science. Without comprehension of meaning, one cannot break down a continuous Vedic text into its constituent words, nor can one know the proper ritual application for such a text. He criticizes the reciters of Vedic texts who do not care for meaning. He refers to Kautsa, who argued that the Vedas are devoid of meaning. Yāska refutes Kautsa's arguments and proposes a number of methods for etymology:

> Now we shall deal with etymology. In this context, words whose accent and grammatical form are regular and have a transparent modification of the formative elements should be explained in the regular manner. However, [for other words,] where the meaning is unclear and the modifications of the formative elements are not transparent, one should always analyse them by focusing on their meaning, by the analogy of some shared action [with another word]. If no such analogy is found, one should explain [words] even by the commonality of a single syllable or sound; but one should never give up an attempt at etymology.
>
> (*Nirukta* 2.1)

While the etymologists were busy offering a breakdown of difficult Vedic expressions, they were clearly aware of a different tradition, namely the tradition of grammar (*vyākaraṇa*). In Yāska's terminology, this reverse process, called *saṃskāra*, consists of building up words by fusing their formative elements and putting them through a series of modifications (*vikāra*). This is the pattern followed in the famous ancient grammar of Pāṇini. It extensively uses the term *artha*, 'meaning', and yet its goal is not interpretation *per se*. It is an encoding grammar rather than a decoding grammar. It begins with the meaning to be communicated and with morphological primes such as roots, stems and affixes. Through a series of steps involving affixation, substitution, modification, and so on, it builds up a surface expression in Sanskrit that corresponds to the intended meaning.

However, Pāṇini's successors, Kātyāyana and Patañjali, clearly see the value of his grammar as a tool for the preservation and comprehension of the Vedic scriptures. They argue that the usage of Sanskrit backed by an understanding of grammar leads to success in life here and hereafter. The grammarians are dealing with a stage of Sanskrit when it is no longer a first language, and needs to be acquired by a deliberate study of grammar. Among the purposes of the study of Sanskrit grammar, Patañjali includes several which relate to one's ability to interpret the scriptural texts properly. One of these purposes is *asandeha*, 'removal of doubts'. Besides mentioning the old *indraśatrú* example, Patañjali cites one more case, *sthūlapṛṣatī*. This word describes a cow that is supposed to be sacrificed. If it is accented as *sthūlapṛṣatī́*, then it can be interpreted as a Karmadhāraya compound to mean 'a cow that is fat and spotted'. However, if it is accented as *sthūlápṛṣatī*, then it becomes a Bahuvrīhi compound and means 'a cow that has fat spots'. A person who does not know the accents or what the difference in accent leads to will not be able to interpret the ritual prescription properly. For Patañjali, studying the Vedas without comprehending their meaning is as futile as throwing firewood away from fire. How can it possibly catch fire?

4 Mīmāṃsā principles of interpretation

The tradition of Mīmāṃsā systematizes the rules of interpretation as they apply to the ritual use of Vedic texts, as well as to the performance of ritual as understood from those texts. Here the term 'Veda' applies to both the mantras and the prose Brāhmaṇa commentaries. For Mīmāṃsā, the Vedas are entirely uncreated, not authored either by humans or by God, yet are fully meaningful and authoritative (see MīmāṃsāČsect;3). Given this eternal

innate authority of the Vedic texts, Mīmāṃsāprovides a set of six hierarchical principles of interpretation, the earlier ones overriding the later ones: (1) direct statement (*śruti*); (2) word-meaning (*liōga*); (3) syntactical connection (*vākya*); (4) context (*prakaraṇa*); (5) position in the text (*sthāna*); (6) name (*samākhyā*).

Let us consider the conflict between the first two guiding principles. Verse 8.51.7 of the *Ṛg Veda* describes the divinity Indra. On the basis of the word-meaning (*liōga*), one would think that this verse should be used while making an offering to Indra. However, a direct statement (*śruti*) in the *Taittirīya-Saṃhitā* (1.5.8), another Vedic text, says that one makes an offering to Gārhapatya fire by using a verse addressed to Indra. This direct statement overrides what one would otherwise decide on the basis of the meaning of the particular verse.

Besides these six principles, Mīmāṃsāalso deals with numerous important issues relating to Vedic interpretation, such as cases where one finds seemingly contradictory statements in the Vedic corpus. How, for example, does one deal with a statement authorizing a Niṣāda chief to perform a certain sacrifice? The tribal people referred to by the term 'Niṣāda' are generally beyond the pale of those eligible to perform sacrifices to Vedic gods. If the word 'Niṣāda-chief' is taken to mean 'one who is a Niṣāda and a chief', this authorization conflicts with the general prohibition of low-caste people taking part in Vedic sacrifices. However, if it is taken to mean 'a chief of the Niṣāda', it could possibly apply to a person who is not a Niṣāda, yet is their chief. The Mīmāṃsakas try to resolve such issues. Where there is ultimately no way to choose between contradictory Vedic injunctions, Mīmāṃsāadvises an optional choice, for example, between 'One uses the Ṣoḍaśīcup in the Atirātra sacrifice' and 'One does not use the Ṣoḍaśīcup in the Atirātra sacrifice'. Considering the entire Vedic corpus to be fully authoritative leads to rejection of its historical and geographical diversity. This rejection gives rise to many more instances of apparently contradictory statements.

5 Other philosophical traditions

Acceptance of the entire Vedic corpus as authoritative requires the reconciliation of a large number of apparently contradictory statements. However, the problem is not unique to Mīmāṃsā. Bādarāyaṇa, in his *Brahmasūtra*, uses the term *samanvaya* ('coordination', 'reconciliation') for the technique of reconciliation. While Mīmāṃsādeals largely with ritual injunctions, Bādarāyaṇa is concerned with the apparently divergent views regarding the ultimate reality of Brahman in the Upaniṣads. The commentators on the *Brahmasūtra* are faced with the same dilemmas, which get more complicated with every generation. Beginning with Śaōkara, the tradition of Vedānta adopts the view that it is founded on three authoritative sources (*prasthānatrayī*), namely the Upaniṣads, the *Brahmasūtra* and the *Bhagavad Gītā*. The principle of *samanvaya* needs to be expanded to all these texts. With Rāmānuja and Madhva, the Purāṇas and the texts of the Pāñcarātra tradition are also added to this pool of authoritative texts. Commentators often create a deliberate hierarchy among different texts, or among different passages from the same text. Then they conveniently argue that some of these texts are to be taken literally (*mukhyārtha*), while others are to be interpreted metaphorically (*lakṣaṇā*). Another dichotomy among texts is stipulated by suggesting that different texts or passages are meant for different audiences, and that these audiences, having different intellectual and spiritual abilities, need suitably different teachings. Some teachings are meant for those who are truly capable, while others are meant to be provisional teachings for those who have not yet arrived at the same high level. Such distinctions are found also in the Jaina and Buddhist traditions (for example, the *upāya/upeya*, *vyāvahārika/pāramārthika* and *neyārtha/nītārtha* distinctions, all of which correspond respectively to 'provisional' and 'true'). With such tools, one can extract a uniformity of teaching from texts which apparently do not have it. The tradition of Dharmaśāstra, religious law, attempted to eliminate similar apparent conflicts between various texts by saying that some were restricted to certain periods of time, regions or social classes.

6 Principles of interpretation in Hindu law

Dharmaśāstra provides an important set of hierarchical guidelines for interpration. Authorities include the Vedas, the Smṛtis (law books), the behaviour of the elites in the community, and, finally, one's own judgment. According to this tradition, and most of the wider Indian tradition, the authority of the Vedic texts overrides that of the Smṛtis. While the Vedic texts are either uncreated or created by God, the Smṛtis are said to be authored by human scholars and are believed to be based on the Vedic texts. The behaviour of the elite leaders of the community is presumed to be based on the Vedas and the Smṛtis. When there is no access to any other authority, one has recourse to one's own conscience (see Duty and virtue, Indian conceptions of §2).

In fact, some early Dharmaśāstra texts say that there were no prescriptions in the Vedas (*śrutyabhāvāt*) regarding the laws that govern different regional, caste and family traditions, and, for that reason, the ancient lawgiver Manu explained these topics. However, along with the tradition of Mīmāṃsā, Dharmaśāstra argues that this simply means that the Vedic texts dealing with such topics are lost, and not that they did not exist. The existence of lost Vedic texts (*anumitaśruti*) is inferred on the basis of statements in the Smṛtis. In general, while the Dharmaśāstra theory lays great emphasis on the Vedas as the ultimate basis for authority, in reality the interpretation of the Vedas gets pulled in the direction of contemporary beliefs and practices.

7 Generalized principles of textual interpretation

Coming back to the wider context of textual interpretation, we find a number of important initiatives. In the tradition of Sanskrit grammar and in the traditions relating to Brahmanical ritual and law, there is a large literature called Paribhāṣā, 'Maxims of Interpretation' (see Abhyankar 1968). In the context of rules of Sanskrit grammar, these maxims help account for the derivation of certain forms, by effectively extending or narrowing the scope of a rule through interpretation. Generally there is great reluctance to alter the wording of a rule or offer a new rule. The tradition would rather get a new ruling by reinterpreting an old one. In doing this, the commentators have recourse to a great many interpretive techniques (see Kielhorn 1887). Related to the Paribhāṣāliterature is the device called *nyāya*. In this context, the term refers to a proverbial statement of colloquial (*laukika*) or technical (*śāstrīya*) wisdom. For example, when two events have no genuine relationship, but are merely coincidental, one cites the *kākatālīyanyāya*, which refers to a story of coincidences, such as the following: a person is resting under a palm tree; a crow flying over the tree comes and sits on a palm fruit; the weight of the crow causes the fruit to fall and crack the person's head. There are a large number of these traditional *nyāyas* (see Kane 1977).

The fifth-century grammarian Bhartṛhari, in his *Vākyapadīya (On Sentences and Words)*, provides a catalogue of many guiding principles of interpretation:

> Meanings of words are differentiated on the basis of the sentence [in which they occur], the context, the meaning [of other words in the context], propriety, time and place [of utterance], and not merely on the basis of the form of a word. Connection, separation, association, opposition, meaning, context, indication, presence of another word, suitability, propriety, place, time, gender, and accent, etc. are the factors which help to determine a specific meaning, in the absence of [natural] clarity of meaning.
>
> (*Vākyapadīya* 2.314-16)

Many of these ideas are carried over into the tradition of Sanskrit poetics, which aims at explaining how one understands the aesthetically pleasing significance of poetry. The notion of interpretation is crucial in all discussions in this tradition. While the lexical meaning of words may be taken for granted, there is a wide gap between lexical meaning and intended meaning. In discussing the phrase 'the cowherd colony is on the river Ganges', we are told that we move away from the literal meaning because of the difficulty of construing the literal meaning (*anvayānupapatti*) and the difficulty of justifying the literal interpretation in view of the intention of the speaker (*tātparyānupapatti*). Nudged by these two factors, we move to the meaning 'the cowherd colony is on the bank of the river Ganges'. Beyond this comes the level of implied or suggested meaning, which may or may not match the meaning intended by the speaker. A listener often understands far more than what the speaker intends. Such levels of suggested and implied meanings are further accounted for through the suggestive function of language and/or inference.

See also: Language, Indian theories of; Meaning, Indian theories of

MADHAV M. DESHPANDE

References and further reading

Abhyankar, K.V. (ed.) (1968) *Paribhāṣāsaṃgraha: a Collection of Paribhāṣā (Texts from Various Systems of Sanskrit Grammar*, Poona: Bhandarkar Oriental Research Institute.(A comprehensive collection of interpretive rules in different schools of Sanskrit grammar.)

Bhartṛhari (5th century) *Vākyapadīya (On Sentences and Words)*, ed. W. Rau, *Abhandlungen für die Kunde des Morgenlands* 42 (4), 1977.(A critical edition of the verse portion of the *Vākyapadīya*; there is no English translation of this, but see Iyer 1969.)

Deshpande, M.M. (1992) *The Meaning of Nouns: Semantic Theory in Classical and Medieval India*, Dordrecht:

Kluwer Academic Publishers.(The bulk of the book is an annotated translation of a seventeenth-century Sanskrit text on the meaning of nouns. The introduction covers the history of a number of semantic theories in Sanskrit grammar and philosophy.)

Gächter, O. (1983) *Hermeneutics and Language in Pūrvamīmāṃsā*, Delhi: Motilal Banarsidass.(An important modern analysis of Mīmāṃsā (theories of language and interpretation.)

Iyer, K.A.S. (1969) *Bhartṛhari: A Study of the Vākyapadīya in the Light of the Ancient Commentaries*, Poona: Deccan College.(A very comprehensive study, though somewhat dated.)

Kane, P.V. (1962) 'On Mīmāṃsā (Principles and Rules of Interpretation in Relation to Dharmaśāstra', in *History of Dharmaśāstra*, Poona: Bhandarkar Oriental Research Institute, 2nd edn, 1977, vol. 5, part 2: 1283-1351.(A seminal study relating the interpretive rules of Mīmāṃsā (to Hindu law.)

Kielhorn, F. (1887) 'Notes on the Mahābhāṣya, No. 7, Some Devices of Indian Grammarians', *The Indian Antiquary* 16: 244-52.(A famous study of the principles of interpretation as used by Sanskrit grammarians.)

Raja, K. (1963) *Indian Theories of Meaning*, Adyar Library Series 91, Madras: Adyar Library and Research Centre.(A somewhat dated, but still very useful, account.)

Yāska (*c*.500 BC) *Nirukta (Analytical Exegesis)*, ed. and trans. L. Sarup, *The Nighaṇṭu and the Nirukta: the Oldest Indian Treatise on Etymology, Philology and Semantics*, London: Oxford University Press, 1920.(Critical edition and translation of an important work attributed to Yāska.)

Language and gender

How do language and gender interact? This can be interpreted as asking about sexual difference in relation to language-use. How do the sexes speak, how do we speak of the sexes? And could or should these patterns change?

Not surprisingly, understanding language-gender interactions solely in terms of sexual difference yields a static and polarized picture. Men insult and swear, women flatter and wheedle, women draw others out while men monopolize conversations, men are direct and women beat around the bush, women gossip whereas men lecture. Linguistic conventions and familiar vocabulary equate humanity with males (note, for example, so-called generic uses of 'he') and sexuality with females ('hussy', for instance, once meant 'housewife'). Men are linguistically represented as actors and women as acted upon, passive. Men control the institutions controlling language - such as schools, churches, publications, legislatures. Children of both sexes, however, learn a 'mother tongue' at a mother's knee.

Such generalizations contain a few grains of truth, at least if restricted to so-called mainstream contemporary America or England. But they completely obscure the differences among women and among men and the varied forms of social relations so important to gender. One is never just a woman or a man: sexual classifications are inflected by age, class, race and much else. And gender involves not only women in relation to men as a group but also more specific cross-sex and same-sex relations ranging from egalitarian heterosexual marriages and same-sex partnerships through intense friendships and enmities among adolescent schoolgirls to camaraderie among boys on a football team. All such relations are partly constituted by people using language to and of one another; all are informed by and inform larger social arrangements. On the more linguistic side, these include dictionaries, the language arts curriculum and editorial guidelines; arrangements with a gender focus include marriage, high-school dances and gay rights legislation.

Emphasizing large-scale sex difference ignores cross-cultural and historical variation and makes change in language, in gender, or in their interaction appear mysterious. And such an emphasis erases the linguistic dynamics of a particular society's construction of gender. Yet it is in such dynamics that, for example, language shapes and is shaped by sexual polarization and male dominance. This entry highlights approaches to language and gender that root each in historically situated social practice. Linguistic change and gender change then become inseparable.

1 Recent background

In the early 1970s, feminist-inspired interest in sex and gender revived venerable philosophical and linguistic questions about how language, culture and thought interact. During the 1980s and 1990s, feminist scholarship in the USA and Europe took an increasingly linguistic turn, not only in philosophy and linguistics but also in history, literary theory, psychology, politics and sociology. From many different viewpoints, scholars argue that language - and more generally, discourse (see §3) - shapes gender identities and relations and supports male and heterosexual privilege. Outside the academy as well, language figures prominently in gender discussions. Mass circulation publications report that the sexes speak 'different languages'; feature writers complain about feminist 'word police' (and often two sentences later about 'misuses' or 'abuses' of words like *rape*); anti-sexist and gay rights activists invent new terminology (*sexism* or *homophobia*) or renovate the old (using *queer* in positive self-reference or *she* as a generic pronoun).

Language-gender debates, however, include participants with different conceptions both of language and of gender. This section touches on three recent kinds of work: the Anglo-American empirically oriented tradition, psychoanalytic theorizing from French philosophers and linguists, and work on discourse and gender construction from feminist philosophers and other theorists in literary and cultural studies.

Linguist Robin Lakoff's widely read *Language and Woman's Place* (1975) framed much subsequent discussion by American linguists and social scientists and some philosophers. Lakoff argued that women face a double-bind. To sound feminine they must speak indirectly and euphemistically; that style, however, is derided as ineffective in public arenas. She argued also that women were systematically derogated when spoken of, that language used of women conventionally implied that they were less worthy and important than men. Although she did not join in

the widespread critique of masculine generics (for example, *man* for humans generally or *he* with a general antecedent; see the introduction to Frank and Treichler (1989) for a discussion of this topic), she noted instructive examples of asymmetries in linguistic resources for speaking of men and of women: *cleaning lady*, for instance, versus the non-occurring *garbage gentleman* or the metaphorical extension of animal or food terms to refer to women, such as *cow* or *tart*.

Others challenged Lakoff's ideas. Linguists such as Janet Holmes (1995) pointed to the multiple functions of linguistic forms. For example, rising intonation on a statement, a form Lakoff interpreted as signalling women's uncertainty or insecurity, can facilitate effective communicative interaction, inviting others to speak. Tag questions ('he's pathetic, *isn't he*?') have a similar range of functions. Lakoff's critics also noted the cultural specificity of speech styles (for example, many African-American women reported Lakoff's characterizations inapplicable to their everyday speech). Furthermore, some (most notably, Penelope Brown (1980) and Marjorie Harness Goodwin (1990)) argued that sex-differentiated rhetorical strategies are not arbitrary cultural conventions. They arise as strategic responses to general social constraints and the specific demands of particular communicative contexts. Women's linguistic agency began to be explored, not just their victimization.

During the same period, some French thinkers began to articulate feminist perspectives on post-structuralist or postmodern views of language. The post-structuralist stance does not take language as a closed system for representing a pre-existing reality (see Postmodernism §2). To speak or write is never a neutral act of encoding, as dominant Anglo-American views of language seem to suggest, but always enacts a self; language is constitutive of subjectivity.

Psychoanalysis, Freud's 'talking cure', has always given language a prominent place, and French psychoanalyst Jacques Lacan made it central in his interpretation of Freud. Linguistic communication, he proposed, is an attempt to erase the first and most traumatic psychic pain, the wrenching separation of birth. To speak is to try to reconnect with the mother, or, more generally, to connect to another; this attempt at (re)connection is inherently phallic. Moreover, to speak and be understood requires one to submit to patriarchal social laws, to place oneself under the Law of the Father. None of this is a matter of being male or female: women can and do speak and write but linguistic communication remains phallic, not feminine, at a deep psychosexual level (see Feminism and psychoanalysis).

Then is silence the only feminine linguistic move? The linguistic and social *status quo*, though powerful, are not stable closed systems. Breaks in the discourse, slips of the tongue, puns, parody and repetition, all bespeak psychic difference and potential resistance, and offer a glimpse beyond the Law of the Father. Such moves disturb and destabilize the patriarchal *status quo*; they can render audible a feminine voice outside the phallic linguistic order. Again it is not sex that is at issue: a man's speech or writing may inscribe a de-centring feminine voice, exposing the instability of the ruling order and threatening its hegemony. Although the French postmodern turn does not necessarily promote the dismantling of male privilege, there is a strong feminist strain in such thinking. The feminine and masculine within each person - a kind of bisexuality - is the fulcrum on which Luce Irigaray and Hélène Cixous, for example, balance femin*ist* politics. Although their specific projects differ, both make femin*ine* outside voices more prominent, and they subvert in varied ways established patriarchal assumptions and values.

Both early American emphasis on sex-differentiated modes of speaking and sexist linguistic resources, and the French focus on the psychosexual significance of language, seemed to take sexual difference and male dominance as given, as prior to linguistic practices. Women and men, the feminine and the masculine - these were unequal poles existing outside language. During the mid-1980s and early 1990s, however, a number of English-speaking feminist philosophers and other cultural and social theorists explicitly questioned the 'naturalness' of heterosexuality and of polarized and hierarchized gender oppositions. They reread thinkers like Irigaray (1977) and emphasized the role of discourse in constructing gender identities and relations, noting that linguistic abstractions and other symbolic practices are central to constructing categories of sex and sexuality. Queer theory especially emphasized the instability of those categories, their emergence from performance, and thus the possibility of alternative forms of gender and sexuality (see, for example, Butler 1990). In a related move, thinkers with a more empirical bent began to conceptualize gender as emerging from social practice (see, for example, Connell 1987).

2 Gender and language in social practice

Much early empirical work took gender and language each as independent structured systems. Researchers asked whether these systems were correlated or connected but not how such connections might arise or why they might matter. Social practice theory suggested a different strategy. Do not look at language and gender as macrosystems; look instead at the microlevel of social activities and conventions for understanding and regulating them. Both gender and language systems can be seen as rooted in social practice; because a single activity often has both gender and language implications, these roots intertwine. Even shared roots do not, however, guarantee that the separate structures supported will connect at higher levels.

What does it mean to say that language is rooted in social practice? Natural language grammars have been argued (most notably by Chomsky and followers) to be tightly constrained by the nature of human minds, seen in this respect as sexless - and likewise unaffected by race, class and geographical origins (see Chomsky, N.; Language, innateness of). From this perspective, fruitful for much linguistic inquiry, social life seems inconsequential to the forms of language. Specifying grammatical possibilities, however, does not tell us everything about language in human history.

At the simplest level, people aim at coordinating their grammars so that communication works reliably within a community. They also talk like those around them simply in order to mark themselves as a social collectivity. Thus controversial questions arise: which grammar(s) or language(s)/dialect(s) should a particular community use? Who adjusts to whom? How do male-dominated institutions affect choice among linguistic possibilities? And (nearer the periphery from a grammatical viewpoint but central for language-users) which words with which meanings are part of shared resources? What difference can vocabulary make to thought and social interaction?

This latter question moves us beyond the form of a community's linguistic resources to the functions those resources serve. What are norms for using language? Are women supposed to listen to men more than men to women? Is it mostly men's linguistic formulations of ideas that enter common currency and serve as background when others speak and write? Who jokes and when? Who swears and when? Who is polite to whom and why? Who suggests and who orders? In which situations and to which addressees? Which meanings are contested? When, where and for whom is silence or talk prescribed? In most communities, doing things with words is highly inflected by gender relations - and also by class, race, age and other dimensions of social difference and hierarchy.

Even more than language, gender is assumed to be built on a biological foundation. Virtually all cultures sort people at birth into one of two sex categories - female and male. External genitalia, which generally predict potential reproductive roles, mark sex at birth. Thereafter, however, other socially constructed markers take precedence. In most social groups sexual classification is associated with power relations, division of labour and regulation of many other areas of life including erotic activity (heterosexism and racism, for example, dictate that desire should be directed only towards someone of the other sex and the same race). Even what we think of as most 'biological' - the roles played in species reproduction - can be affected by social activity. Consider such historical developments as bottle-feeding, frozen sperm and test-tube babies. Just as strikingly, cultures differ dramatically in the range of gendered identities they offer (not all, for instance, stop with two sexes) and the kinds of gender relations that prevail.

The substantive content of gender derives from social practices and the attitudes and expectations that drive and support them. Some areas of gender practice seem independent of language: for example, styles of dress and other kinds of bodily adornment and demeanour, conventions regulating who touches whom when and where and what the reaction should be, participation in competitive sports, food preparation and serving. Even these, however, connect to linguistic practices. A person's clothes and visual style may get them dubbed 'slutty' or 'elegant', silence is deemed consent to sexually charged touch, locker-room talk is judged unsuitable for female ears, a group of women talks in the kitchen while their male partners trade comments about the football game on the living room television.

Writing is also important. Teen magazines have articles on lip gloss and diet, self-help books advise on female orgasm, newspapers devote much of their space to sports stories and use sports metaphors even in articles about business or politics, women's but not men's magazines include many pieces on food for family and for entertaining. Of course neither speech nor writing is just words. Teen magazines, for example, link to huge cosmetics and fashion industries and to the myriad of institutions and practices pushing teenage girls towards

(exclusively heterosexual) romance. But the bottom line is that social practices of all kinds have rich linguistic texture: language is central to social life.

3 Connecting two different conceptions of discourse

The emphasis of much recent feminist thinking has been on discourse as a global feature of culture, what I will call cultural discourse. We might, for instance, talk about the (cultural) discourse of romance in which talk, writing, clothing, photographs, film and much else present romance as normatively heterosexual, females' primary aspiration in life, the only important form of cross-sex relations, and so on. Cultural discourse covers background assumptions, favoured rhetorical strategies, vocabulary, and nonlinguistic activities and representations that highlight some perspectives on a particular domain (for example, intimate interpersonal relations) and obscure others.

Linguists and many philosophers of language use a notion of discourse more narrowly linguistic and less global. What I will call a situated discourse consists of a historically located series of connected utterances or inscriptions - for example, a conversation or a written narrative. For analytic purposes we can identify a (situated) discourse with a string of sentences together with relevant aspects of the contexts in which the sentences are produced and interpreted. Cultural discourse can be seen as grounded in situated discourses, whose detailed analysis may shed light on the microstructure of larger discursive constructions of gender.

Studies of language in use give concrete linguistic content to claims about the discursive production of sex, sexuality and gender. For example, names and styles of address start forming sexed identities early on (nurses in one hospital were heard calling baby girls 'sweetheart' or 'beautiful', but using 'Jones' for the baby boy with that surname). Adolescents draw on categorizing practices that link sex and sexual decorum with social class when they label someone 'slut'. Engagement in same-sex verbal tussles (direct insults, often overlaid with humour, or 'he-said/she-said' tales about absent parties), heterosexual harassment (street comments on a woman's body) or the discourse of romance (identifying four-year-olds' opposite-sex playmates as their 'sweethearts') - such everyday language-use helps shape and sustain particular forms of gender identity and relations.

Less everyday language-use is also, of course, critically important. Sermons or scriptures may enjoin women's silence in places of worship or other forms of sex-differentiation and hierarchy. And social scientists write about gender, often equating it with statistically significant sex differences (and ignoring its inseparability from matters of class, race, gender and even age). We find, for instance, sweeping and dichotomizing claims about how women and men speak. Such work is popularized and becomes enormously influential. It gets used not only for interpreting but for regulating gender-language connections: not only is this 'on average' how women and men speak but this is how 'real' or 'nondeviant' women and men speak or ought to speak. (Although Tannen (1990) herself avoids such normativizing, her many readers have not.)

Social life depends on social interaction, whether face-to-face or more diffuse and larger in scale. Interactions are made up of more than situated discourses, of course, but language is central to most of them and virtually essential to large-scale and long-distance social exchange. Situated discourses are the primary ingredient of social life: cultural discourses derive much of their substance from situated discourses.

All situated discourses are constrained by conventions and institutions predating them. All, however, have potential effects on subsequent discourses and, more generally, on social structural constraints affecting talk and understanding, on cultural discourses and other macro-level social arrangements. For example, when talking about people in general, saying 'Each must do what she thinks best' will startle many listeners: English classes have prescribed 'Each must do what he thinks best' in such contexts. The feminine generic goes against familiar conventions, yet its increasing use (notably in American philosophical writing) begins to establish an alternative convention. The more it is used the less it shocks.

4 Changing language, changing gender

'Sticks and stones may break my bones but words will never hurt me', chant children trying to defuse the sting of labels like 'sissy', 'faggot' or 'bitch'. They seek to convince themselves and their tormentors of the view, widely endorsed in one form or another, that language has no force, is causally inert. As formal objects on their own, of course, words (more generally, languages) do not hurt people or do anything else to them. But words in use do

indeed affect people in many ways: they convince, persuade, enlighten, frighten, humiliate, amuse, disgust, titillate. Words link to social arrangements on the one hand (see Speech acts), and to causally relevant features of courses of events on the other (see Reference). These links give them many kinds of power in human affairs and thinking, both individual and collective. Thus it is not surprising that language change and social change might go hand in hand.

Take a simple example. Labelling people affects how those people and their labellers then enter into a host of social practices. Until recently most communities using American-English conventionally assigned adult women social titles on the basis of marital status ('Mrs' or 'Miss'), whereas men's social titles were not so differentiated ('Mr' being the only option other than occupational and professional titles). The introduction of 'Ms' offered women a title option supposedly neutral as to marital status. This option was adopted by a diverse group including advertisers who did not want to offend by making mistaken assumptions about marital status, young unmarried women wishing not to advertise themselves as single, and self-described feminist women claiming a status equivalent to that of male peers (whether husbands or not). There have been other changes in linguistic practice connected to marriage, sexuality and family: for example, more women retain a pre-marriage surname or hyphenate surnames, some couples create a new shared surname, and some use the mother's surname for a child; even women who have adopted a husband's surname frequently favour 'Ms/Mrs Jane Doe' over the formerly dominant 'Mrs John Doe' form; employers and others now inquire about a 'spouse' (or even a 'partner', finessing both marital status and sexual preference) rather than a 'wife'; marriage ceremonies much less often ask a woman to 'obey' a husband, and those officiating frequently pronounce the couple 'husband and wife' rather than 'man and wife'; same-sex couples go through marriage-like rituals and adopt common surnames; 'parent' has acquired a use as a verb (which is semantically far closer to the verb 'to mother' than to the verb 'to father').

Such changes in sociolinguistic practice have accompanied nonlinguistic changes in the institutions of marriage and family and practices associated with them. Middle-class women are less likely than they used to be to assume their economic welfare and social position will derive from husbands' income and status. Not only do young women see divorce as a real possibility, they also see themselves as capable of significant earnings and professional achievement, whether or not they opt to marry or to have children. Some men see caring for children - parenting - as a central role for part of their lives. Heterosexual marriage of the traditional hierarchical and strongly sex-differentiated kind, though still in many ways the default option for middle-class Americans, is increasingly seen as not the only choice open, not even for women who want to bear children. Adding 'Ms' as a title option for women and changing laws about surnames of married women and their children could not have increased women's participation in high-level careers if all else had remained the same, just as having available the sex-neutral verb 'to parent' does not suffice to get men more actively engaged in the activity it denotes. But changes in linguistic practice have been (and continue to be) part and parcel of changes in gender practices - the linguistic and nonlinguistic developments reinforce one another.

Linguistic practices change all the time. For example, new words or terms are introduced. 'Surrogate mother' designates a kind of relation of woman and child not earlier envisaged (and carries with it certain assumptions about the social weightiness of that relation); 'sexual harassment' groups together, on the basis of similar effects, kinds of situations previously either ignored or seen as very different in kind ('sexual teasing', for instance, and 'seduction'). Of course, new expressions draw on associations with existing ones: 'sexism', for example, developed meaning in part through the implied analogy with the word 'racism'. Sometimes an existing form is altered in its uses, as when 'partner' comes to designate the person with whom one lives in a sexually intimate relationship, whether that relationship is sanctioned by the state or religious authorities, whether the person so designated is a woman or a man, and whether of the same sex as, or opposite sex to, oneself. But the implication of equality remains from other uses of the term 'partner'.

Customary usage may change while reference stays fixed. Some have begun to say 'my child' or 'my kid' in contexts where most people still say 'my daughter' or 'my son', to say 'kids' rather than 'girls and boys', and in other ways to resist identifying everyone always in sex-specific terms. The linguistic system and its interpretation are not thereby changed, but patterns of language-use are and with them what is implicated when people speak, what we take them to mean above and beyond what their words literally say (see Implicature). There are no explicit norms that say adults must identify a person to children by using 'that woman' or 'that man' or similar sex-specific forms, yet this is standard practice: 'Say thank you to the nice lady' rather than 'Say thank you to the

nice person who gave you that sweet'. Changes in such practices might ultimately help effect major shifts in gender polarization and emphasis on sexual difference, an emphasis important for enforcing heterosexuality.

Could some group impose what might seem desirable linguistic and related social changes? Should they? History shows that some regulation is possible: for more than a century, schools and editors, for example, have proscribed singular 'they' (used by Shakespeare, Jane Austen and many others, including me and the authors of other *Routledge Encyclopedia of Philosophy* entries) and prescribed a supposedly sex-indefinite 'he'. Simply removing institutional sanctions (poor grades, having an article rejected) from the use of singular 'they' would almost certainly greatly reduce use of supposedly generic 'he'. But of course some people might continue its use, and even those who have dropped it might nonetheless speak in other ways indicating a view of humans as normatively male: 'During the night the villagers left in canoes, leaving us behind with the women and children.'

Linguistic conventions, both those that narrow the range of linguistic systems on which community members can draw and expect to be understood and those that promote certain patterns of usage (for instance, when to say 'thank you'), only constrain and never completely determine what community members will say. People can in various ways challenge and resist such conventions or exploit indeterminacy in them as to form and meaning. And other people can make countermoves as in the charges of a silly 'political correctness' hurled at those who have drawn attention to social biases implicit in existing linguistic conventions and have proposed alternatives. No social standpoint monopolizes all moves; subordinated interests can find expression. This does not mean that social advantage confers no linguistic advantage (it does) or that linguistically aided social change is not possible (it is). But total control of language use (and therefore of the patterns of thought and action it might facilitate) is only an Orwellian nightmare, not a real possibility. And, as many feminist thinkers have reminded us, quick linguistic fixes for sexism and heterosexism are just pipedreams (see the guidelines in Frank and Treichler 1989).

See also: Feminism; Linguistic discrimination

SALLY McCONNELL-GINET

References and further reading

Brown, P. (1980) 'How and Why are Women More Polite: Some Evidence from a Mayan Community', in S. McConnell-Ginet, R.A. Borker and N. Furman (eds) *Women and Language in Literature and Society*, New York: Praeger, 1980, 111-36.(Draws on Grice-style pragmatic theory and her own ethnographic work to develop a strategic view of politeness and, more generally, speaker style. Reprinted in Coates 1997.)

Butler, J. (1990) *Gender Trouble*, London and New York: Routledge.(Develops her view of the performativity of gender and its base in discourse.)

Cameron, D. (1985) *Feminism and Linguistic Theory*, London: Macmillan, 2nd edn, 1992.(Argues that linguists' limited conceptions of language and their inattention to feminist theory have impeded their understanding of language and gender.)

Cameron, D. (ed.) (1990) *The Feminist Critique of Language: A Reader*, London and New York: Routledge. (Extracts from a variety of sources, both literary and social scientific, including Lakoff's early writing and an interview with Irigaray.)

Coates, J. (1996) *Women Friends Talking*, Oxford: Blackwell.(Situated discourses and their detailed analysis.)

Coates, J. (ed.) (1997) *Language and Gender: A Reader*, Oxford: Blackwell.(Reprints many of the hard-to-get 'classics' on the topic.)

Connell, R.W. (1987) *Gender and Power*, Stanford, CA: Stanford University Press.(A social practice theory of gender; see especially chapter 4, 'The Body and Social Practice'.)

Eckert, P. and McConnell-Ginet, S. (1992) 'Think Practically and Look Locally: Language and Gender as Community-based Practice', *Annual Review of Anthropology* 21: 461-90; repr. in C. Roman, S. Juhasz and C. Miller (eds) *The Women and Language Debate: A Sourcebook*, New Brunswick, NJ: Rutgers University Press, 1994.(Develops social practice view of language and of gender discussed in §2.)

Frank, F.W. and Treichler, P.A. (eds) (1989) *Language, Gender and Professional Writing: Theoretical Approaches and Guidelines for Nonsexist Usage*, New York: Modern Language Association.(Includes several articles on how meanings develop in discourse and are sanctioned institutionally, as well as intelligent and detailed guidelines accompanied by discussion of the complexities of language-use.)

Goodwin, M.H. (1990) *He-Said-She-Said: Talk as Social Organization among Black Children*, Bloomington, IN: Indiana University Press.(Draws on ethnomethodological theory to argue that social structure is constituted in large part by the conversational structures community members produce in their daily interactions, examples of

which are analysed in detail.)

Hall, K. and Bucholtz, M. (eds) (1995) *Gender Articulated: Arrangements of Language and the Socially Constructed Self*, London and New York: Routledge. (Papers are all written for this volume and include Lakoff's recent work, an excellent anthropological essay by S. Gal, an essay by Eckert and McConnell-Ginet developing and applying the theoretical framework introduced in their 1992 article, and many other diverse and detailed studies of language-use.)

Holmes, J. (1995) *Women, Men and Politeness*, London: Longman.(Draws on empirical studies she and others have done of politeness and gender in language-use, noting especially multiple functions of linguistic forms.)

Irigaray, L. (1977) *Ce Sexe Qui N'En Est Pas Un*, Paris: Éditions de Minuit; trans. C. Porter with C. Burke, *This Sex Which is Not One*, Ithaca, NY: Cornell University Press, 1985.(The essay by the same name gives a short and relatively accessible introduction to her thinking.)

Lakoff, R. (1975) *Language and Woman's Place*, New York: Harper & Row.(An influential argument that women are victimized by language.)

McConnell-Ginet, S., Borker, R.A. and Furman, N. (eds) (1980) *Women and Language in Literature and Society*, New York: Praeger; repr. New York: Greenwood, 1994.(First collection to bring together empirical and theoretical work from social sciences, literature and philosophy: see especially article on defining 'anger' by N. Scheman.)

Roman, C., Juhasz, S. and Miller, C. (eds) (1994) *The Women and Language Debate: A Sourcebook*, New Brunswick, NJ: Rutgers University Press.(Collects a number of papers from literary, psychoanalytic and social scientific authors.)

Tannen, D. (1990) *You Just Don't Understand: Women and Men in Conversation*, New York: William Morrow. (An anecdotal best-seller focusing on supposed sex differences in communicative styles.)

Vetterling-Braggin, M. (ed.) (1981) *Sexist Language: A Modern Philosophical Analysis*, Totowa, NJ: Littlefield, Adams & Co.(This collection contains many philosophical discussions of language-gender issues, including a debate on the ethics of 'Ms', along with more linguistic contributions, such as V. Valian's critique of Lakoff's failure to distinguish a language system from its uses.)

Wright, E. (ed.) (1992) *Feminism and Psychoanalysis: A Critical Dictionary*, Oxford: Blackwell.(See especially the entry 'Language', by J.F. MacCannell, and its accompanying bibliography.)

Language of thought

The 'language of thought' is a formal language that is postulated to be encoded in the brains of intelligent creatures as a vehicle for their thought. It is an open question whether it resembles any 'natural' language spoken *by anyone. Indeed, it could well be encoded in the brains of people who claim not to 'think in words', or even by intelligent creatures (for example, chimpanzees) incapable of* speaking *any language at all. Its chief function is to be a medium of* representation *over which the* computations *posited by cognitive psychologists are defined. Its language-like structure is thought to afford the best explanation of such facts about animals as the productivity, systematicity and (hyper-)intensionality of their thought, the promiscuity of their attitudes, and their ability to reason in familiar deductive, inductive and practical ways.*

1 The language of thought hypothesis (LOTH)

One of Descartes' greatest challenges to the materialist theory of the mind was that it was impossible to imagine a physical device capable of the diverse intelligent behaviour often exhibited by human beings. Frege's development of formal logic, and then Turing's conception of a Turing machine offered the promise of a serious reply to this challenge, since they showed how at least deductive reasoning could be realized as a form of mechanically realizable computation (see Turing machines; Mind, computational theories of). Their proposals were an inspiration for the development of computational formalisms for various forms of non-deductive reasoning: induction, abduction, practical reason and decision theory.

But computations presuppose representations (Fodor 1975): the computations of Turing machines and many computers are, for example, often defined over *numerals* that represent *numbers*. A suggestion that has been advanced by a variety of philosophers and cognitive scientists (Harman 1973; Newell and Simon 1972; Fodor 1975, 1987) is that thinking requires precisely the sort of representations that are standardly used to *identify* thoughts, namely *sentences*: representations with logico-syntactic structure, of the sort defined recursively in logic texts in terms of names, predicates ('is bald'), variables ('x', 'y'), connectives ('and', 'only if'), quantifiers ('all', 'some') and various operators ('probably', 'necessarily'). Thus, the thought that 'Necessarily God does or does not exist' is normally identified by using that sentence (or a translation of it).

The language of thought hypothesis (LOTH) is then the hypothesis that thinking consists in performing computations on sentences whose logico-syntactic parts are causally efficacious, for example, by being encoded in a creature's brain. The process is akin to theorem-proving in logic, except that, where the rules of logic are ordinarily applied by us *consciously following* the rules, according to LOTH the rules are applied by virtue of the causal structure of the brain. For example, where, in elementary logic, we follow the rule *modus ponens* - 'From "p" and "If p then q" derive "q"' - for LOTH the brain is so constructed that, if it is in a state that represents the premises, then it is (sometimes) caused to enter a state that represents the conclusion.

'Thinking' here is a generic word for processes involving propositional attitudes, and these are distinguishable in at least two ways that are often blurred in ordinary talk: by their *contents* (the sentence complement, for example, the that-clause, that follows the attitude verb) and by the different *relations* the agent may have to the same such contents, for example, *believing* or *hoping* that God exists (see Propositional attitudes). According to LOTH, different propositional attitudes of an agent involve different computational relations borne by the agent to sentences in a language of thought (LOT) that express the various thought contents of the agent. To a first approximation (see also Field 1981):

For any agent, x, and propositional attitude, A that p, there exists some computationally definable relation C_A such that:
x A's that p iff   for some σ: $(xC_A\sigma \& \sigma)$ means that p

For example, actively judging or desiring that God exists might be defined in terms of different computational relations, *J* or *D*, that an agent, *x*, might bear to a sentence, '$(\exists x)Gx$', that expresses the proposition 'God exists'.

LOTH must, of course, specify the computational relations for specific attitudes, as well as provide an account of how a symbol in the brain can have a specific meaning. For the first task it appeals to the familiar flow-charts of cognitive psychology. Judgment, for example, might be the output of perceptual and reasoning systems that is the

input to a decision-making one. For a theory of meaning, LOTH has turned to (combinations of) 'informational' approaches (Fodor 1987), teleological approaches (Neander 1995) and 'computational role' approaches (Field 1981).

Note that, *pace* Searle (1984; see also Chinese room argument), LOTH does not entail that syntactically defined sentences in a LOT do not have many real and important semantic properties; all that is claimed is that the relational clause and the causal processes of thinking are specifiable syntactically. However, there are some LOTH proponents (for example, Jackendoff 1987) who do seem to think that syntax would be enough to determine semantics, and others (for example, Stich 1983) who argue that the semantic clauses are not in principle determinate.

2 Arguments for LOTH

Although some philosophers (for example, Davies 1991) have ventured a priori arguments for LOTH, the chief arguments for it are that it provides the best explanation of at least the following phenomena (see Rey 1997: chaps 8-9 for a longer list).

The productivity of attitudes. People seem to be able to think a potential infinitude of thoughts, that is, to a first approximation, people can (in principle) think all permissible combinations of the primitive syntactic elements. For example, they can understand a conjunction of $n + 1$ sentences if they can understand *n* of them. Some theorists have baulked at the substantial idealization (from memory, mortality) that this involves, and so Fodor (1987) proposed a related, but more modest claim:

The systematicity of attitudes. Anyone who can think *p* can think any logical permutation of *p*: for example, if someone can think that 'Ann hates Bob only if Charles loves Di', they can also think that 'Charles loves Di only if Ann hates Bob', 'Di loves Charles only if Bob hates Ann' and so on for all permissible logical permutations. LOTH captures both productivity and systematicity by presuming that any system in which the logico-syntactic elements of sentences are causally efficacious is one in which they are readily available for recombination.

Rational and irrational relations among attitudes. Both deductive reasoning and many common fallacies are 'structure-sensitive', involving the scopes of operators, for example, negations, conditionals and quantifiers (see Scope). Quantifier scope is what distinguishes, for example, 'Everyone loves someone' form 'There is someone whom everyone loves'. Standard treatments of logic capture such structural facts in the terms of logical syntax. By insisting that that syntax is causally efficacious, LOTH is able to explain people's ability to reason, and why they are prone to certain errors (for example, misrepresenting a scope).

The (hyper-)intensionality of attitudes. Propositional-attitude ascriptions are 'intensional': terms that (even necessarily) refer to the very same thing cannot be substituted for one another without risking a change in the truth-value of the whole. There is a difference between thinking 'Water is wet', thinking 'H_2O is wet', and thinking 'The stuff of rain is wet', despite the fact that water = H_2O = the stuff of rain. LOTH distinguishes these attitudes by distinguishing syntactically between different symbolic structures to which an agent can be related (it can even distinguish them when the structures have the same 'meaning', as in *hyper*-intensional cases such as remembering that a fortnight is a fortnight as compared with remembering that a fortnight is two weeks).

The multiple roles of attitudes. Different attitudes can be directed at the same thoughts. People often *wish* for the *very same thing* that they *believe* does not presently obtain; they often come to *think* what they previously only *feared*. LOTH captures this by positing different computational relations to the same internal representation (or ones with the same content).

3 Some common objections to LOTH

Many people claim introspectively not to 'think in words', but, rather, for example, in mental images (see Imagery). Now, in the first place, LOTH is not meant to be establishable by introspection; it is purely an *explanatory* hypothesis. Moreover, by careful delineation of the different computational roles of specific representations, some (for example, Pylyshyn 1981) have argued that it can capture imagistic experiences, and even subjectivity and sensation (for example, Lycan 1990; Rey 1992).

But, second, it is difficult to think of a non-linguistic representational system with anything like the expressive power of a linguistic one. *Purely* imagistic systems, for example, do not seem adequate to represent logically complex thoughts - for example, negations, conditionals, nested quantifications - nor to distinguish thoughts about a general category (for example, cow) from ones about a particular instance (Elsie), for which the same image might serve.

Note that LOTH is not committed to the language of thought being confined only to creatures that speak a natural language, much less to the language of thought actually *being* a natural language. According to LOTH, any creature that *thinks* (as, for example, chimpanzees seem to do) will need a LOT, whether it *speaks* or not.

Wittgenstinians often object that LOTH presupposes the very sorts of processes it purports to explain: if there is a LOT, do we not need an 'homunculus' in the brain to read it? LOTH answers this objection by emphasizing Turing's proposal for computation in general, whereby brute causation replaces human calculation.

But did Wittgenstein's 'private language' argument not show that languages are necessarily public? (See *Philosophical Investigations* §§243-; Private language argument; Wittgenstein, L. §13.) Wittgenstein was concerned, however, only with languages whose references could not be publicly ascertained (for example, because they referred to 'private sensations'). LOTH makes no such commitment.

Dennett (1987: ch. 3) argues that many attitude ascriptions are merely 'interpretations' of behaviour that do not commit us to a corresponding sentence in the head. For example, we might say of a chess-playing computer that 'It likes to get its queen out early', even though there might be no corresponding sentence manipulated by the computer's program. However, LOTH is not committed to the literal truth of *every* ordinary ascription; only ones that figure in a causal explanation of such phenomena as those mentioned above.

The chief rival to LOTH is the so-called 'radical connectionist' models of cognition (in contrast to connectionist models that are merely implementations of 'classical' ones such as LOTH). Defenders of radical connectionism complain that LOTH models are too rigid, inefficient and not as neurophysiologically realistic as connectionist networks (see Smolensky 1988). However, the cognitively relevant physical properties of the brain have yet to be sufficiently identified to pass judgment on such a claim, and it has yet to be shown that radical connectionist theories can explain all the above phenomena as well as a language of thought does (see Connectionism).

See also: Mind, computational theories of

GEORGES REY

References and further reading

Davies, M. (1991) 'Concepts, Connectionism, and the Language of Thought', in W. Ramsey, S. Stich and W. Rumelhart (eds) *Philosophy and Connectionist Theory*, Hillsdale, NJ: Erlbaum.(An a priori argument for the language of thought.)

Dennett, D.C. (1987) *The Intentional Stance*, Cambridge, MA: MIT Press.(Essays critical of LOTH, presenting a less realistic, more 'instrumental' theory of mental states instead.)

Field, H. (1981) 'Mental Representation', in N. Block (ed.) *Readings in Philosophy of Psychology*, vol. 2, Cambridge, MA: Harvard University Press.(An extremely lucid presentation of the essential claims of LOTH, as well as of a conceptual role semantics for it.)

Fodor, J. (1975) *The Language of Thought*, New York: Crowell.(The most influential defence of LOTH as a psychological hypothesis.)

Fodor, J. (1987) *Psychosemantics*, Cambridge, MA: MIT Press.(A defence of LOTH on the basis of systematicity, as well as a development of a specific version of informational semantics for it.)

Harman, G. (1973) *Thought*, Princeton, NJ: Princeton University Press.(An early statement of a language of thought view, more from a philosophical than a psychological point of view.)

Jackendoff, R. (1987) *Consciousness and the Computational Theory of Mind*, Cambridge, MA: MIT Press.(A linguist's attempt to work out in some detail a LOTH view, incorporating research in vision and language.)

Loewer, B. and Rey, G. (1991) *Meaning in Mind: Fodor and his Critics*, Oxford: Blackwell.(Collection of papers reacting to Fodor's defence of LOTH, with replies by Fodor and an extended introduction by the editors.)

Lycan, W. (1990) 'What is the 'Subjectivity of the Mental'?', in J.E. Tomberlin (ed.) *Philosophical Perspectives*, vol. 4, *Action Theory and Philosophy of Mind*, Atascadero, CA: Ridgeview, 109-30.(An attempt to account for subjective experiences by means of LOTH.)

Maloney, J. (1989) *The Mundane Matter of the Mental Language*, Cambridge: Cambridge University Press.(A clear and very readable defence of LOTH, with replies to many critics and consideration of problems outside its normal range, such as sensations and moods.)

Neander, K. (1995) 'Misrepresenting and Malfunctioning', *Philosophical Studies* 79: 109-41.(An evolutionary (teleosemantic) approach to the problem of assigning content to expressions in a language of thought.)

Newell, A. and Simon, H. (1972) *Human Problem Solving*, Englewood Cliffs, NJ: Prentice Hall.(An influential statement of the 'classical' LOTH approach in artificial intelligence.)

Pylyshyn, Z. (1981) 'The Imagery Debate: Analog Media vs. Tacit Knowledge', in N. Block (ed.) *Imagery*, Cambridge, MA: MIT Press.(Argues that LOTH can capture imagistic experiences.)

Rey, G. (1992) 'Sensational Sentences Switched', *Philosophical Studies* 67: 77-103.(An attempt to account for sensory experience, and the puzzle of inverted qualia, by means of LOTH.)

Rey, G. (1997) *Contemporary Philosophy of Mind: A Contentiously Classical Approach*, Oxford: Blackwell.(An introduction to recent work and an attempt to show that LOTH handles traditional problems in philosophy of mind.)

Searle, J. (1984) *Minds, Brains and Science*, Cambridge, MA: Harvard University Press.(A well-known attack on LOTH.)

Smolensky, P. (1988) 'A Proper Treatment of Connectionism', *Behavioral and Brain Sciences* 11: 1-23. (Influential critique of LOTH and suggestion of a connectionist alternative.)

Stich, S. (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief*, Cambridge, MA: MIT Press.(A clear discussion of LOTH, and a defence of it as a purely syntactic account of thought.)

Language, ancient philosophy of

The earliest interest in language during the ancient Greek period was largely instrumental: presumed facts about language and its features were pressed into service for the purpose of philosophical argumentation. Perhaps inevitably, this activity gave way to the analysis of language for its own sake. Claims, for example, about the relation between the semantic values of general terms and the existence of universals invited independent inquiry into the nature of the meanings of those general terms themselves. Language thus became an object of philosophical inquiry in its own right. Accordingly, philosophers at least from the time of Plato conducted inquiries proper to philosophy of language. They investigated:

(1) how words acquire their semantic values;
(2) how proper names and other singular terms refer;
(3) how words combine to form larger semantic units;
(4) the compositional principles necessary for language understanding;
(5) how sentences, statements, or propositions come to be truth-evaluable;

and, among later figures of the classical period,

(6) how propositions, as abstract, mind- and language-independent entities, are to be (a) characterized in terms of their constituents, (b) related to minds and the natural languages used to express them, and (c) related to the language-independent world.

1 Pre-Platonic figures

Before Plato, evidence for inquiries into the semantic and syntactic features of language is sketchy, if suggestive. Even so, there is clear interest in language among pre-Platonic philosophers, rhetoricians and Sophists. In some cases this interest is motivated by a concern for developing persuasive speech, primarily for forensic or political use (see Rhetoric §1). In other cases, it has a more philosophical orientation. Thus, quite early, Parmenides (early to mid-5th century BC) sought to exploit a putative fact about language when defending an improbable form of monism: he argued that 'what is for saying and thinking must be' (fr. 6) and concluded that it is impossible to speak or think of what does not exist, with the consequence that it is equally impossible to speak or think of generation, change, or plurality, since each of these is implicated, directly or indirectly, in what is not. With the additional thesis that it is possible to speak and think of everything which exists (fr. 3), Parmenides concludes that the universe is an eternal undifferentiated unity, altogether remote from the appearances of plurality, change and generation delivered by the senses (see Parmenides §2). When Plato and Aristotle later exposed the fallacy in this argument, it was mainly by distinguishing syntactic and semantic features of the verb *to be* which Parmenides had failed to mark.

The Sophist Gorgias (late 5th century BC) similarly appealed to features of language when attempting to establish the nihilistic theses that:

(1) nothing exists;
(2) even if something did exist, it would be unknowable;
(3) even if something did exist and were knowable, it would be incommunicable.

On behalf of the odd counterfactual (3), in particular, he appealed to a subjectivist form of meaning: if something were perceived, it could be communicated only by means of a sign, which, even on the untenable assumption that it could be made to represent the supposed object of perception perfectly, could not be shared by more than one mind. Hence, communication about objects perceived is impossible (see Gorgias §3). However implausible this argument may sound, it is worth remarking that it appeals to a thesis about meaning that many find natural and attractive; indeed, many centuries later, Frege still felt the need to mount positive arguments against just its kind of semantic subjectivism (see Frege, G. §4).

The atomist Democritus (mid 5th-4th century BC) reportedly composed a treatise *On Words*. In this or another work he inaugurated an important discussion, taken up with great earnestness by later philosophers, concerning the signification of names. Democritus thus stakes out what appears to be the first case for a dedicated semantic

theory. He thus moves beyond the primarily instrumental interest in the semantic features of language displayed by his predecessors.

Democritus evidently argued for a form of *conventionalism*, according to which the relation between names and things named is determined by agreement or legislation. He argues that there cannot be a natural relation which ties a word to the world, since:

(1) there are homonyms, that is one name for different entities (for example, 'bank');
(2) there are synonyms, distinct names for the same thing;
(3) entities can change names in the course of their existence.

Consequently, it would be wrong to suppose that a name is determined by anything other than an arbitrary designation.

In arguing this way, Democritus means to reject *naturalism*, a view according to which names and things named are related by some non-conventional relation. A first approximation of naturalism might be the view that all naming is at root onomatopoetic or somehow pictographic. Put thus crudely, naturalism may seem a non-starter. But other forms of naturalism, loose by comparison with these simpler versions, appeal instead to natural descriptions, with the result that names, especially proper names, all turn out to be disguised definite descriptions. Making naturalism more tenuous still, a fair number of naturalists were in the habit of appealing to the origin of names and other words, thereby confusing genetic and semantic questions.

Although in these varying contexts, the varieties of 'naturalism' were unhelpfully conflated, the contrast between conventionalism and naturalism is important for the history of classical semantics for three related reasons. First, the ensuing debate, much of it quite subtle and sophisticated, is initially couched in terms of a contrast between conventionalism and naturalism, and even in late antiquity data to which naturalists appeal appear in a surprisingly forceful way. Second, part of the motivation for naturalism turns out to be reasonable: naturalism, unlike some varieties of conventionalism, seemed to capture the normativity of language. Some names are correct and fitting, while others are incorrect or otherwise inappropriate. It is hard to specify precisely how this can be the case if spoken sounds are applied to things in a purely conventional manner, especially when conventions may be generated quite locally. Third, whatever its ultimate faults, naturalism at least attempts to explain how reference occurs. In some of its ancient forms, conventionalism had difficulty even broaching a suggestion on this point.

2 Plato

The opposition between conventionalism and naturalism dominates Plato's most explicit and extended treatment of language, the *Cratylus* (see Plato §15). The dialogue opposes equally untenable naturalist and conventionalist theories of naming, while at the same time promoting the virtues of each. Plato focuses initially on proper names, but it is clear that his interest extends to other singular terms and to general terms as well. Perhaps Plato means to explore facts about reference by laying out extreme forms of views which, though incompatible, each have some merit. At any rate, in the course of the dialogue he sets aside the most implausible features of conventionalism and naturalism without detailed comment.

These implausible features are, however, fully present in the initial presentations of each theory. Cratylus opens the dialogue, arguing on behalf of naturalism that it is possible for someone not to be named *n* even if 'all people were to call' him *n* (383b6-7, cf. 429b12-c2). This is because, as we later learn, '*n*' is the name of something only if it bears a natural relation *R* to that thing, where *R* is specified by Cratylus only incompletely as a kind of mimetic or imitative relation (423b-428a, 430a10-b1). Thus, something qualifies as a name *n* only if it is correct, that is, only if it imitates the thing named and indeed indicates what that thing is (428e1-2).

Hermogenes responds on behalf of conventionalism:

> I cannot be persuaded that there is some correctness (*orthotēs*) of names other than convention and agreement. For it seems to me that whatever name someone gives to something, this is the correct one…. For no name is suited by nature to anything, but rather by the custom and habit of those in the habit of using it and calling things by that name.

> (*Cratylus*, (384c10-d9)

Hermogenes' conventionalism may seem initially more plausible than any variety of naturalism. Yet he is taken to task for failing to consider external constraints on convention. More importantly, names have the dual function of conveying information and distinguishing real natures in the world 388b7-11); if a name misses or obscures these real natures, then it will not succeed as a name, or at any rate will be at best only a sub-optimal name. Plato suggests that even this degree of normativity is incompatible with the version of conventionalism Hermogenes espouses.

In opposing this kind of conventionalism, Plato does not adopt Cratylus' naturalism. On the contrary, the Socrates of the dialogue, who evidently represents Plato's point of view, is equally critical of a naturalism which fails to distinguish between the correct name of an entity and the name which people actually employ when designating that entity. Socrates thus twice embarrasses Cratylus by forcing him to allow that his theory cannot distinguish between successfully referring to an object and referring to it correctly (429c3-5 and 429e8-430a5). The point, updating Plato's example a bit, is this: if a child points to a whale, exclaiming, 'That fish is bigger than our house!', the child has evidently successfully referred to the whale, but only incorrectly. Cratylus must allow that not only has the child failed to designate the whale, but also may have produced a false sentence by referring to a minnow in the whale's neighbourhood.

Plato, in the end, agrees with the naturalist that there must be some relation R which in a direct or indirect way relates a name to a thing named; but he denies that R can be specified in a way which wholly ignores convention. So, he equally agrees with the conventionalist that convention is relevant to determining names; but he denies that convention by itself can determine R. For he believes that names are information-bearers which, if correct, reflect the structure of the extra-linguistic world. Thus, the argument of the *Cratylus* has the effect of distilling what is right from both conventionalism and naturalism, and of laying down the constraints for a semantically and metaphysically adequate account of reference. It also advances the debate by distinguishing different ways in which a name may be said to be correct or incorrect: a name may be correct in a pragmatic way by fulfilling its reference-fixing function, or correct in a descriptive way by carrying information which accurately reflects the world. Plato appreciates that these functions may come apart in ways which make the job of specifying R exceedingly difficult. He himself does not, however, offer an articulated account of the no doubt complex relation R. Instead, the dialogue ends in perplexity.

Although it contains his most self-conscious and sustained treatment, the *Cratylus* hardly exhausts Plato's views on naming or language more generally. Of special note are his complex discussions of the various syntactic and semantic functions of the verb *to be*, conducted most thoroughly in the *Parmenides* and *Sophist*, both late dialogues. In the (*Parmenides* (142a, 161e-162b), raising a puzzle about saying truly 'a is not' (we cannot predicate something of *a* unless *a* is; but then we cannot do what we plainly can do, namely deny the existence of something), Plato takes up a discussion about negative existentials which continues even today (see Existence). When, in the *Sophist*, he responds to the Parmenidean argument for monism, Plato quite rightly focuses on distinct uses of 'is'. The distinction remains an important philosophical tool for recognizing and diagnosing fallacies.

3 Aristotle

Aristotle approaches the study of language with a logician's eye. He seeks to determine how terms relate in syntactic structures of various sorts, mainly to determine and regiment the correct and incorrect forms of inference-drawing. Accordingly, he investigates the nature of meaning, the difficulties of reference, and the errors which result from failing to attend to what he calls the *homonymous* uses of words, that is the uses of words with a plurality of distinct but connected meanings, including especially core philosophical terms whose non-univocity may initially elude us. He thinks that when they seek unified definitions of these sorts of terms - at any rate, non-disjunctive definitions given as necessary and sufficient conditions - philosophers ignore the complexity of the concepts or properties they are used to express. Natural language sometimes reflects this complexity; at other times, it obscures it.

Aristotle investigates linguistic phenomena primarily in the *Categories* and *De Interpretatione*, but also in the *Metaphysics*, the *Topics*, the *Sophistical Refutations*, and the *Prior* and *Posterior Analytics*. He does not always explicitly connect these investigations, but there are several bridge passages that indicate how he takes his views in these areas to be related. Perhaps the single most important such passage opens the *De Interpretatione*, where Aristotle draws a fairly complete analogy between words and sentences on the one hand and types of affections

(*pathēmata*) in the soul on the other.

According to this analogy, there is a reasonably straightforward relation between thoughts and spoken sounds: individual words are to assertoric sentences as individual thoughts are to compound thoughts. The first members of these pairs are without truth value, being in some sense semantically atomic. The second members are by contrast necessarily either true or false. That is, Aristotle claims that bivalence obtains for all simple assertoric sentences and their mental analogues, compound thoughts.

Partly in virtue of this analogy, the opening of the *De Interpetatione* contains the seeds of four related Aristotelian semantic theses, each of which is developed in various places in the corpus:

(1) *Compositionality.* The semantic value of assertoric sentences is a function of their sub-sentential semantically relevant parts (semantically relevant, because, as he points out *De Interpretatione* 16a20-22), neither 'ton' nor 'dent' contributes anything to the meaning of 'Clinton was unusually well educated for an American president.');
(2) *Conventionality.* The written and spoken symbols used to stand for thoughts are conventional, whereas the thoughts themselves and that for which they stand are not;
(3) *Relationalism.* Conventional semantic units (written marks and spoken sounds) receive their semantic significance from those things of which they are symbols;
(4) *Signification.* The relationship in virtue of which they receive their semantic significance consists in or involves what Aristotle calls signification (*sēmainein*).

Of these, (2) and (4) merit special attention.

The second thesis, conventionalism, indicates where Aristotle stands on the question explored in the *Cratylus*. He explicates his contention thus:

> I say <a name is a significant sound> according to convention (*kata sunthēkēn*) because no name is by nature <significant>, but only when it has become a symbol. Even though inarticulate noises, e.g. those belonging to wild beasts, do reveal something (*dēlousi ti*), none of them is a name.
>
> (*De Interpretatione* 16a26-29)

He rejects naturalism, but without endorsing any simple version of conventionalism. Instead, he lays down the constraint that a sound becomes a name only when it has become a symbol, where something is a symbol only when it stands in an appropriate relation to a non-conventional mental representation.

With respect to Aristotle's approach to language and meaning, the fourth thesis, regarding signification (*sēmainein*), is easily the most central and important. It is also the most complex and difficult. Aristotle often writes as if signification were a simple meaning-relation: words signify things, even when they lack referents. Thus, for example, after distinguishing sharply between assertoric expressions, which alone have truth-value, and their constituents, which are not yet truth-evaluable, he imagines someone objecting that the word 'goatstag' (*tragelaphos*) is already false, since there are no such creatures. He responds that although 'goatstag' signifies something, it is not yet true or false, precisely because it is not yet part of an assertoric sentence (*De Interpretatione* 16a16-18). Aristotle's response correctly drives a wedge between vacuous reference and falsity; it also seems to treat signification as closely akin to sense expression: 'goatstag' has sense, but lacks reference (see Sense and reference).

Still, a complication arises for two related reasons:

(1) Aristotle sometimes appeals to signification where it seems unlikely that it can be understood in terms of sense expression;
(2) Aristotle sometimes denies signification to vacuous singular terms.

In the first case, Aristotle surely maintains that not only words signify: the word 'man' signifies rational animal, but the entity man signifies rational animal as well (*Categories* 3b10-23; *Topics* 122b16-17, 142b27-29; *Posterior Analytics* 85b18-21; *Metaphysics* 1017a22-27, 1028a10-16); further, clouds signify rain and smoke signifies fire (*Posterior Analytics* 70a10-38).

Elsewhere, he denies that a single word made to mean 'manandhorse' has signification (*De Interpretatione*

18a19-27). The probable resolution is that the signification relation is for Aristotle in some contexts semantic and in some contexts not; further, in some contexts, signification is more akin to reference and in others more akin to sense expression. In some more technical contexts, Aristotle introduces signification as a sort of essence-specification: man signifies rational animal because humans are essentially rational. Perhaps in this sense, though, 'signification' behaves all the more like the English word 'meaning': smoke means fire; 'ghoul' means 'a spirit that robs graves and devours the corpses'; and we say that although he speaks of piety, Euthyphro does not even know the meaning of the term. In some philosophical contexts, we are appropriately technical about how we understand the concept of meaning; in some other contexts, we are comparatively relaxed. This practice mirrors Aristotle's treatment of signification (see Semiotics).

4 The Stoics and other Hellenistic movements

The Hellenistic schools which flourished after the death of Aristotle became increasingly technical and specialized in their treatments of language. This is especially true of the Stoics, whose interests in logic, grammar and syntax led them to offer deeply subtle theories replete with technical vocabularies capable of an unprecedented richness and precision. The most intricate and important innovations were most likely introduced by Chrysippus.

It is difficult to recapitulate elements of the Stoic system briefly. To begin, the evidence is fragmentary, deriving from many different sources, some of them hostile and most of them secondhand. Moreover, there is a fair amount of divergence within the Stoic camp itself; it is worth remembering that we are dealing with the data of several centuries of philosophy. That said, the doxographer Diogenes Laertius provides a serviceable overview of the main features of Stoic semantic theory:

> Utterance (*phōnē*) and speech (*lexis*) differ, because while vocal sound is also an utterance only articulate sound is speech (*lexis*). And speech differs from language (*logos*), because language is always significant (*sēmantikos*) whereas speech <can> also lack significance, for example, 'blityri'; language can in no way <lack significance>. Moreover, saying (*to legein*) differs from voicing (*to propheresthai*). For while utterances are voiced, what is said are states of affairs - which turn out to be things said (or things which can be said, *lekta*).
>
> (Diogenes Laertius 7.57)

The initial view here is straightforward. Animals, including humans, make noises; but only some of these noises are meaningful. Those noises which express *lekta* are meaningful, or *sēmantikos*, and those which do not are mere sounds. If I speak German and you do not, the utterance '*verkehrt*' will be meaningful for me, but not for you; but even if we both speak German, the nonsense utterance 'blityri' will be meaningless for us both. The Stoics suggest, then, because some noises are significant and others are not, we must suppose that some noises have meanings, namely express *lekta*, while others do not. Hence, it is necessary to postulate the existence of *lekta*.

The primary motivation for the introduction of *lekta* is, then, semantic. For better or worse, with the Stoics we have the first self-conscious reification of meanings as such. Unsurprisingly, the Stoics were quick to exploit a second semantic function of *lekta*. Some *lekta*, called by the Stoics *axiōmata*, are the principal bearers of truth and falsity. According to the Stoics, then, one species of *axiōmata* can be true or false, evidently in a primary way. Our speaking truly or falsely, that is our uttering sentences with truth values, depends upon our uttering a sentence which expresses a complex of meanings, an *axiōma*. Hence, the sentence 'Dion is walking' is true just in case it expresses the proposition, held by the Stoics to be noncorporeal, *that Dion is walking* and that proposition is true. Just as words are correlated with meanings or significates, which the Stoics regard as incomplete *lekta*, so the complete sentence 'Dion is walking' gains its meaning and truth-evaluability by expressing the complete *lekton*, the *axiōma* or proposition which is true independent of its being expressed. Just how determinate Stoic thinking about propositions is remains disputed, since some reports have the Stoics introducing propositions which change truth-values and which, though noncorporeal, can perish, or go out of existence. Often enough, though, in these cases the Stoics turn out to be grappling with quite subtle problems generated by indexicals and demonstratives. Indeed, they often show more sensitivity to these problems than their critics and later expositors display when discussing their views (see Stoicism §8).

The Stoics are by no means alone in offering important advances in thought about language during the Hellenistic period. Sextus Empiricus certainly develops semantic themes. Additionally, the Epicureans, and to a lesser extent the Academics, offer treatments of meaning and language understanding, often reverting to earlier debates about

natural versus conventional signification. Later, the eclectic physician and philosopher Galen (AD 129-*c*.210) returns to the naturalism-conventionalism debate, ridiculing the naturalist propensity for etymology as 'fine friend' which is also an 'impostor'. No other Hellenistic school rivals the Stoics, however, in their genuinely innovative and impressively technical handling of semantic and syntactic matters.

See also: Dialectical school; Language, Indian theories of; Language, philosophy of; Nature and convention; Proper names; Propositions, sentences and statements; Reference

CHRISTOPHER SHIELDS

References and further reading

Annas, J. (1982) 'Knowledge and Language: the *Theaetetus* and the *Cratylus*', in M. Schofield and M. Nussbaum (eds) *Language and Logos*, Cambridge: Cambridge University Press.(An exploration of some overlooked connections between Plato's approach to naming in the *Cratylus* and the difficult dream passage of the *Theaetetus*.)

Aristotle (*c.* mid 4th century BC) *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984, 2 vols.(Although the focus of several works, Aristotle's attention to semantic matters pervades much of the corpus.)

Atherton, C. (1993) *The Stoics on Ambiguity*, Cambridge: Cambridge University Press.(A comprehensive treatment of Stoic approaches to ambiguity, with a special emphasis on Stoic classifications of forms of ambiguity and their role in fallacious argumentation.)

Bolton, R. (1976) 'Essentialism and Semantic Theory in Aristotle', *Philosophical Review* 85: 514-44.(An investigation into Aristotle's approach to nominal definition and the semantics of natural-kind terms.)

Brown, L. (1986) 'Being in the *Sophist*, a Syntactical Enquiry', *Oxford Studies in Ancient Philosophy* 4: 49-70.(A clear and accessible investigation into Plato's treatment of being in the *Sophist* which challenges some orthodox interpretations.)

Everson, S. (ed.) (1994) *Companions to Ancient Thought 3: Language*, Cambridge: Cambridge University Press. (An extremely useful anthology of papers on many aspects of ancient philosophical thought about language, with a special emphasis on continuities between ancient and contemporary concerns. This work should be consulted for additional bibliography.)

Fine, G. (1977) 'Plato on Naming', *Philosophical Quarterly* 27: 289-301.(An exceptionally clear and forthright account of Plato's approach to naming.)

Irwin, T. (1982) 'Aristotle's Concept of Signification', in M. Schofield and M. Nussbaum (eds) *Language and Logos*, Cambridge: Cambridge University Press.(A challenge to the widely held view that signification in Aristotle is a kind of meaning-relation; instead, Irwin argues, signification is a technical relation of essence-specification.)

Kretzmann, N. (1971) 'Plato on the Correctness of Names', *American Philosophical Quarterly* 8: 126-38.(A detailed and highly nuanced investigation of Plato's *Cratylus*.)

Long, A. and Sedley, D. (1987) *The Hellenistic Philosophers*, Cambridge: Cambridge University Press, 2 vols. (The first volume contains translations of the principal fragments of Hellenistic philosophy, together with commentary. The second volume contains original texts and a very useful bibliography.)

Nuchelmans, G. (1973) *Theories of the Proposition: Ancient and Medieval Conceptions of the Bearers of Truth and Falsity*, Amsterdam: North Holland.(A full assessment of the evidence for the introduction of propositions in ancient thought.)

Plato (390s-347 BC) *Plato: Complete Works*, ed. J.M. Cooper, Indianopolis, IN: Hackett Publishing Company, 1997.(This edition contains the best translation of the *Cratylus*, together with the rest of Plato's works, many of which contain sub-sections relevant to his views on language.)

Shields, C. (1998) *Order in Multiplicity: Homonymy in the Philosophy of Aristotle*, Oxford: Oxford University Press.(A full account of Aristotle's approach to homonymy, discussed within the broader context of his semantic theory.)

Language, conventionality of

When we say that smoke means fire or that those spots mean measles, we are noting how the presence of one thing indicates the presence of another. For these natural relationships to continue, it is enough that the laws of nature remain the same. The connection between the two states is strictly causal. By contrast when we say, 'In English, "gold" means this stuff", pointing at some metal, we are insisting on an arbitrary connection between a piece of language and part of the world. We might have used another word, as other languages do, or have used this word for something else. But, for a word to have the literal meaning it does in a language, this arbitrary connection must be sustained on subsequent occasions of use. What is needed to sustain the connection is an intention on our part, not just the continued operation of natural laws.

Of course, some connections between words and things are based on natural relations; there is, for example, onomatopoeia. However, few words have this feature. For the majority of words it is quite arbitrary that they have the meanings they do, and this has led many to suppose that the regularities needed to sustain the connections between words and what they stand for are conventional rather than causal. But there are also those who deny that convention is an essential feature of language.

1 Language as governed by conventions

Words do not have intrinsic meanings. They mean what they do because speakers have given them this meaning (see Meaning and communication). But how do they do this? At first we might focus on what a speaker means by their use of words on a given occasion. This may be fixed by their communicative intentions. But for words to have a literal meaning in a public language they must have the same meaning from one occasion to the next, and many have supposed that this can be established only by conventional practice. Conventions, it is claimed, establish the regularities that give words the meanings they have for groups of speakers. So while a speaker may succeed in communicating something by their utterance of an expression on a given occasion, its literal meaning among a population of speakers is what it is conventionally used to communicate on each occasion of use.

Among those who think the appeal to conventions is essential in the study of language, there is room for disagreement about the precise explanatory role they play. Conventions are arbitrary practices that can be defined nonlinguistically; and for some this provides a reason to think they can serve in an analysis of the concept of literal meaning. The analysis proceeds in two stages. If we can explain how a speaker succeeds in meaning something by their use of words on a particular occasion then we can try to explain how this is established as the conventional meaning of those words in the common language of a given population.

An example of this programme is provided by Stephen Schiffer (1972). Schiffer begins with Grice's analysis of speaker-meaning in terms of the speaker's beliefs and intentions to explain what an expression means on a single occasion of use (see Grice, H.P.; Communication and intention). He then argues that we must appeal to conventions to show how speaker-meaning can give rise to the regular and literal meaning of expressions in a public language. A sequence of sounds or marks will have the literal meaning it does among a group of speakers when there is a convention to use those sounds or marks with a particular speaker-meaning. It will no longer be necessary to work out the speaker's meaning each time; the prevailing conventions among a population of speakers will make certain sounds and marks meaningful to them but not necessarily to others, and languages will be patterns of conventional regularities sustained by the practices of particular groups.

A different explanatory role for conventions is proposed by David Lewis (1975). Conventions governing use do not constitute language or determine meaning but they do establish which language is the actual language used by a population. For Lewis, *languages* are well-defined abstract objects - sets of expression-meaning pairs - which exist independently of speakers. We can characterize a language as a function which maps strings of signs to meanings. Among populations of humans there will exist conventions of *language* (in a second and more general sense of the term). These will be regularities in verbal behaviour depending on the beliefs and actions of people in those populations. These conventional regularities determine which of these abstract objects is the actual language used by that population. A population of speakers can be said to use a given language *L* just in case the meaning an expression has in *L* is the meaning it is conventionally taken to have in that population. The account is not reductive, however, for it assumes that speakers already know the meaning of sentences in a given language and

choose to conform to these. But it does offer an analysis of the conventions needed for members of a population to speak the same language. This goes via Lewis' general analysis of conventions which runs as follows.

A regularity *R* is a convention in a population *P* if and only if:

(1) Everyone in *P* conforms to *R*.
(2) Everyone in *P* believes the others are conforming to *R*.
(3) The belief that the others are conforming to *R* gives each person in *P* a good reason to conform to *R*.
(4) People in *P* prefer general conformity to *R* to less than general conformity to it.
(5) *R* is not the only regularity it would be possible for everyone in *P* to conform to.
(6) Conditions (1) to (5) are mutually known to members of *P*.

On this model it is in our common interest to conform to the regularities and abide by the conventions of using a given language. This facilitates cooperation and secures communication. We can coordinate our activities by communicating reliable information to one another by using the same words. This gives us a motive for wanting the regularities to be upheld.

The actual language relation is specified as follows. For a language *L* to be the actual language used in *P* is for there to be a convention of truthfulness and trust in *L*; that is, a convention of using particular sentences in *L* to make statements only if members of *P* believe them to be true, and to take statements other people make in *L* to be a basis for forming beliefs about what is true. To do this speakers must use sentences in accordance with their meanings in *L* just so long as other members of *P* do so. In this way Lewis seeks to reconcile the formal study of languages with activities of particular populations of human language-users.

Followers of Wittgenstein would reject both of the above accounts, criticizing as mysterious the idea of conventions correlating expressions with meanings *qua* entities, and rejecting the idea that meaning in a speech community can be explained in terms of a prior account of what individuals mean by their words on particular occasions. Once we accept that the meaning of expressions is determined by communal use, no appeal to Platonic entities is necessary. But since the rules that govern linguistic practice are determined by the community as a whole, an individual's understanding of language must be parasitic on communal practice. In contrast to the attempted reduction of literal meaning to conventions and speaker-meaning, an individual can only use language meaningfully when participating in the shared practices of a linguistic community. Thus Michael Dummett (1991) thinks that in using words individual language-users must hold themselves responsible to the standards of use of the language to which those words belong. On this view, a language is constituted by the conventional practices and agreed standards of usage. In addition, Dummett believes there are overarching conventions governing the overall aim of language-users to aim at the truth, as well as conventions governing the force (for example, assertoric, interrogative, imperative, and so on) with which they make their utterances.

Dummett offers an argument for the view that language must be governed by communal conventions. It concerns use of terms where we rely on experts to determine their precise meaning. The claim is that I can use the words 'fullback' or 'leveraged buyout' meaningfully while knowing little about football or finance because I rely on there being others who can use these words with greater precision. This is the phenomenon Hilary Putnam (1975) calls the division of linguistic labour (see PUTNAM, H. §3). It depends on a speaker making use of a word that has both an everyday and a more technical meaning, deferring to experts to fix the precise application of the term. However, this feature of language, although pervasive, is not essential to the existence of meaningful speech and we could get by without it. It does not show that there must be conventions of use among, or deference to, a set of experts for me to use words meaningfully.

2 Problems with convention-based accounts

In a revision to his earlier thinking, Schiffer (1987) argues that the reductive analysis of literal meaning is unlikely to succeed since every account of the notion of speaker-meaning we have devised fails. He makes powerful criticisms of Lewis' account of the actual language relation. Speakers can understand sentences of their language they have never heard before: there are potentially infinitely many of these. Clearly, there are no conventions governing the 'still to be used' part of a language. So unless the part we use fixes the unused part there will be nothing to determine which language is the actual language of a group of speakers. It is implausible that speakers

know a function which maps an infinity of sentences onto meanings. But perhaps what they know about the conventionally governed part of the language provides the resources to use and understand the rest. This would require a convention-based account of the expression-meaning of subsentential items, and the rules for their combination. Yet it is equally unlikely that speakers knowingly subscribe to conventions governing each word and grammatical construction ever used. A more plausible suggestion is that speakers have internalised a grammar of the language and this provides the semantic and syntactic resources to comprehend any sentence belonging to it. If this is correct, then why should we think it is conventions rather than psychological mechanisms that relate speakers to their languages? Schiffer argues that just as many difficulties confront us in attempting to relate speakers to grammars as to particular languages.

3 Language without conventions

Objections in principle have been raised to the idea that conventions have an essential role to play in the analysis of language. Donald Davidson and Noam Chomsky have both denied that language requires the existence of conventions. Davidson (1986) rejects the idea that conventions are necessary for linguistic competence. Chomsky (*1986*) rejects the idea that they are sufficient. For Davidson, Lewis-style conventions are too restrictive to explain successful communication. For Chomsky, they are insufficiently systematic to explain a speaker's knowledge of language. (This was the basis of the criticism in §2.)

Both Chomsky and Davidson have opposed the idea of language as a social practice bound by a shared set of rules or conventions. Instead they argue that the fundamental notion of language is that of an individual's language, or idiolect. For Chomsky, language has no existence outside the mind of the individual language-user. It is a body of knowledge represented in the mind/brain of the speaker. Davidson, on the other hand, views language as necessarily social, requiring the existence of others with whom we communicate. Despite their differences, neither accepts that we need to share a language in order to communicate.

For Davidson, meaning something by one's words depends on one's intention to be understood in a particular way and whether one succeeds. What we mean depends on what we can be understood to mean. This is possible even when speaker and hearer do not share the same meaning for a word, as the case of malapropism makes clear. Davidson goes further in conjecturing that there is no reason in principle why speakers who understand each other have to share any aspects of their language. What they would rely upon instead are general principles for interpreting one another's behaviour: principles for ascribing meanings to each other's words, so as to make rational sense of someone's utterance given the beliefs and desires that best explain that person's behaviour. This is not to deny that there are conventional aspects to language use, or that in actual speech situations people will make use of similar syntactic devices and enjoy overlapping vocabularies. But none of this is essential for communication. Even where conventions do exist in a language, it is possible to depart from them and still be understood. How far we can depart cannot be determined in advance nor settled by another set of conventions. Thus, Davidson argues, we need neither appeal to conventions nor stick to them to be understood, so they cannot play an essential role in linguistic understanding.

Chomsky agrees with Davidson that for a speaker and hearer to understand one another they do not need to agree on the meaning of words or abide by shared conventions. He departs from Davidson in rejecting the idea that success in communication is a condition of significant language use. Communication is never guaranteed and it does not enter into the conditions for something's being a language. For Chomsky, language is determined by the cognitive psychological facts internal to the speaker's language faculty. No two speakers will share the same idiolect; each will have their own internalized grammar and lexicon. If anything is shared it is the initial state of the language faculty characterized by the principles of universal grammar. For Chomsky, universal grammar, which comprises a set of structural principles true of all human languages, is a genetic endowment (see Language, innateness of). Speakers have no conscious or explicit knowledge of it, so cannot choose to conform to it. Nor, if Chomsky is right, is there any alternative to relying on it. Its operation cannot count, therefore, as a convention of language according to Lewis' definition. For Chomsky it is universal grammar, and not the existence of conventions, that explains the possibility of human languages.

Whether one accepts Chomsky's view of languages, his empirical findings point to difficulties for Lewis' proposal. For if no two speakers ever share the very same vocabulary items or observe exactly the same rules of grammar there will be no formally precise, syntactic and semantic characterization of the language spoken by a

given community. On the other hand, what the regularity account of conventions correctly stresses is the social and public aspect of linguistic use and meaning, although in insisting on shared practices, or conformity to communal standards, it goes beyond what is necessary to ensure these. Davidson's arguments show that meanings can be established publicly without having to be conventional. Conventions of language may abound but they have not been shown to play an essential role in language.

See also: Language, social nature of

BARRY C. SMITH

References and further reading

Chomsky, N. (1986) *Knowledge of Language*, New York: Praeger.(Chapters 1 and 2 offer a clear exposition of his psychological conception of an individual's language.)

Davidson, D. (1986) *A Nice Derangement of Epitaphs*, in E. Lepore (ed.) *Truth and Interpretation*, Oxford: Blackwell.(Presents the case against conventions and shared public languages.)

Davies, M. (1981) *Meaning, Quantification, Necessity*, London: Routledge.(Chapter 1 presents a clear and detailed account of the analytical programme discussed in §1.)

Dummett, M.A.E. (1991) *The Logical Basis of Metaphysics*, Cambridge, MA: Harvard University Press.(Chapter 4 summarizes Dummett's views about language, arguing against Davidson and Chomsky and in favour of a communal conception of language.)

Lewis, D.K. (1975) *Languages and Language*, in A. Martinich (ed.) *The Philosophy of Language*, Oxford: Oxford University Press, 2nd edn, 1990.(Key paper expanding on material discussed in §1.)

Loar, B. (1981) *Mind and Meaning*, Cambridge: Cambridge University Press.(A well-argued account of the project to define semantic notions in psychological terms with good observations of the difficulties noted in §2 for convention-based accounts of meaning.)

Putnam, H. (1975) 'The Meaning of "Meaning"', *Collected Papers*, vol. 2, Cambridge: Cambridge University Press.(An important paper covering many themes beyond the scope of this entry. Of use here for the account of the division of linguistic labour.)

Schiffer, S. (1972) *Meaning*, Oxford: Oxford University Press.(A densely argued book with elaborate discussion of the analytic programme described in §1. Not elementary.)

Schiffer, S. (1987) *Remnants of Meaning*, Cambridge, MA: MIT Press.(A closely argued rebuttal of his earlier views and of any attempt to provide a theory of meaning. Challenging but not for beginners.)

Schiffer, S. (1993) 'Actual Language Relations', *Philosophical Perspectives* 7: 231-58.(A thorough assessment of attempts to define these relations.)

Language, early modern philosophy of

Philosophical interest in language during the seventeenth and eighteenth centuries was strong but largely derivative. Most thinkers shared Leibniz's view 'that languages are the best mirror of the human mind, and that a precise analysis of the significations of words would tell us more than anything else about the operations of the understanding'.

The three most important areas of philosophical discussion about language in the modern period were the nature of signification, the origin of human language and the possibility of animal language. Signification was generally viewed as a relation between linguistic signs and ideas. There was no agreement whether signification is entirely conventional or contains a natural element, but the view is that it is entirely natural virtually disappeared. Even those who retained the belief in the possibility of a philosophically perfect language insisted that such a language should be constructed anew, rather than rediscovered as the lost language of Adam. The traditional biblical account of the origin of language was more and more contested but, as more naturalistic theories emerged, the problem of why other animals cannot talk became especially pressing.

Debates about language in the seventeenth and eighteenth centuries were highly speculative; participants in these debates often relied on simplistic biological theories, inadequate grammars or anecdotal evidence from travellers. What makes these discussions important is less their scientific contribution than their engagement with the philosophical problems concerning the relationship between the human mind and the natural world.

1 The biblical and Lucretian accounts

Two classical accounts provided the intellectual background for modern philosophical debates about language. The first is the biblical story according to which the original names of beasts and birds were given by Adam. Renaissance speculation contends that these names were inspired by God, bore a natural relation to the creatures, and captured their essences. After the confusion at Babel, Adam's universal language was scattered into mutually unintelligible forms of speech, each of which began to decay. The second influential ancient story comes from Lucretius' *On the Nature of Things* (V.1028-90). According to Lucretius, language is an institution created by a community of humans, rather than a single name-giver. People came to form articulate sounds out of a need to coordinate their actions. At the beginning, the ability to communicate by words differed little from children's ability to indicate the object they desire by means of gestures, or animals' ability to indicate fear or pain by cries. As language developed, its structure became more complicated and its uses more diverse.

Perhaps the most criticized point of the biblical account is the claim that language originates with the naming act of a single person. Ideas from Hobbes' *Leviathan* (1651), Locke's *Second Treatise on Government* (1690) and Rousseau's *Du contrat social (The Social Contract)* (1762) provide the paradigm of how a social account of the origin of language might be framed. According to the contractarian view, social norms - whether they are laws of property or laws of meaning - are the result of explicit or tacit agreements among human beings. There is a striking parallel between Locke's criticism of the natural origin of the title to sovereignty in the *First Treatise on Government* (1690) and his opposition to the natural foundation of categorization in the *An Essay Concerning Human Understanding* (1689). Who is to be the heir is not to be decided through the study of the line of inheritance from Adam; what is to be gold is not to be decided through the study of how the word 'gold' descended from the original naming act of Adam. It is rationally grounded convention that settles these matters. Philosophers of the Enlightenment followed Locke's analysis and brushed away the biblical account. Condillac, Rousseau and Herder argued that even if language happens to be a divine creation, human beings had to reinvent it after the Flood or after they scattered over large uninhabited areas, leaving the biblical story at best irrelevant.

The most controversial point of the Lucretian account is the claim that the difference between the primitive signalling of animals and human languages is only a matter of degree. Descartes argued that the ability to understand an unlimited variety of expressions is distinctive of a mind, something that the smartest animal could never do, but which is no challenge for even the dullest human being. This shows 'not merely that beasts have less reason than men, but that they have no reason at all' (Descartes 1637: 58). In his objections, Gassendi protested against this conclusion, and insisted that animals do in fact speak. The Cartesian response - analysing animal signalling and the bodily aspects of human communication in physiological terms - was accomplished in great

detail by Géraud de Cordemoy. The conclusion was that since such an analysis provides a full account of animal signalling, but not of human language, the latter but not the former presupposes a mind (see Animal language and thought).

2 Empiricism: Hobbes

Although not primarily concerned with philosophy of language, Thomas Hobbes presented an influential and distinctively empiricist account of language in Chapter 4 of *Leviathan* (1651) and Part 1 of his *De corpore (On the Body)* (1655).

According to Hobbes, the purpose of speech is to 'transferre our Mentall Discourse, into Verbal; or the Trayne of our Thoughts, into a Trayne of Words' (1651: I.4). There are two reasons we make such a transfer: to record our thoughts for ourselves, and to communicate our thoughts to others. Words employed as mnemonic devices are *notae* (marks); words employed as means of communication are *signa* (signs) (1655: II.1-2). Hobbes holds that the first use of words is primary: if a person were alone in the world, he could create a private language and use words as marks, but not as signs ((1655: II.3).

Hobbes' philosophy of language has two main parts. The first is the account of the relationship between words and what they stand for, or the theory of *names* 1655: II); the second is the account of the connections names bear to each other, or the theory of *propositions* (1655: III).

Names signify conventionally: Hobbes thinks that even if the original names of animals come from God himself, he nevertheless chose those names arbitrarily (1655: II.4). What a name signifies is a *cogitatio* (thought), for the claim that a word like 'stone' is a sign for stones can only be understood as saying that 'he that hears it collects that he that pronounces it thinks of a stone' (1655: II.5). (Nonetheless, Hobbes often speaks of names as signifying things without the mind.) Hobbes is committed to the empiricist thesis that all human thinking is based on imagination. Since imagination itself is nothing but decaying sense, we cannot meaningfully speak of things that cannot be thought of as combinations of sensory images we had previously. Hence, for example, we cannot think of actual infinity: saying that something is infinite means merely that we are 'not able to conceive the ends, and bounds of the thing named' (1651: III.12).

Hobbes' semantics is purely nominalistic (see NOMINALISM §3). General terms do not pick out properties, but signify disjointedly every individual to which they apply (1655: II.9). In a proposition, both terms 'raise in our mind the thought of one and the same thing', and the copula 'makes us think of the cause for which those names were imposed on that thing' (1655: III.3). The proposition '*man is a living creature* raises but one idea in us, though in that idea we consider that first, for which he is called man, and next that, for which he is called living creature' (1655: V.9).

Conventionalism about signification and nominalism in semantics prepare the ground for conventionalism about truth. Truth is ascribed only to propositions, and it consists in the fact that the predicate is a name of everything that the subject is a name of. According to Hobbes, 'the first truths were arbitrarily made by those that first of all imposed names upon things, or received them from the imposition of others. For it is true (for example) that man is a living creature, but it is for this reason, that it pleased men to impose both those names on the same thing' (1655: III.8). It is controversial whether Hobbes consistently adhered to this radical view (see Hobbes, T. §3).

3 Rationalism: Port-Royal

Grammaire générale et raisonnée (General and Rational Grammar: The Port-Royal Grammar) (1660) and *La logique ou l'art de penser (Logic or the Art of Thinking)* (1662) articulate a rationalist conception of language sharply opposed to Hobbes' empiricism. Following Descartes, Arnauld and his collaborators (Claude Lancelot in *Grammar* and Pierre Nicole in *Logic*) reject not only empiricism, but conventionalism and nominalism as well.

The empiricist claim that all our thoughts originate in perception leaves unexplained how we attain ideas like those of being and thinking. Since - the authors of *Logic* insist - we do have such ideas, it follows that 'the soul has the faculty to form them from itself, although often it is prompted to do so by something striking the senses' (1662: I.1). The element of truth in conventionalism is that the association between a word and an idea is arbitrary. However, since the connections among ideas and between ideas and things are natural, we reason not about words

but 'about the nature of things by considering ideas of the mind that people chose to mark by certain names' (1662: I.1). Finally, while nominalists are right to claim that all existent things are singular, some entities are not things, but rather modes of things. Modes exist only in virtue of being instantiated in things, but they can be thought of independently of the things in which they inhere 1662: I.2, I.6, II.1).

The semantic terminology of *Grammar* and *Logic* is often loose. The authors talk about words signifying (*signifier*), expressing (*exprimer*) or marking (*marquer*) both things and ideas. (The relationship between ideas and things, however, is consistently called *représentation* (representation).) The principal meaning (*signification principale*) of a word or a sentence is the idea or proposition it expresses, but in order to capture its full meaning connotations must also be taken into account. These connotations are called *auxiliary ideas* (*idées accessoires*). For example, the principal meaning of the sentence 'You are lying' is simply 'You know that the contrary of what you say is true', but in addition to the principal meaning, these words also convey the idea of contempt (1662: I.14).

The principal meaning of a word can often be presented by an explicit definition. Following Aristotle's distinction and scholastic terminology, Arnauld and Nicole distinguish between *nominal definition* and *real definition* (1662: I.12). A nominal definition captures a connection between a word and an idea; a real definition captures a connection between ideas, which in turn corresponds to a connection between real entities. Therefore, nominal definitions - at least in their purest form - are arbitrary and are not subject to rational criticism. On the other hand, real definitions in their purest form 'do not depend on us at all, but on what is contained in the true idea of a thing' (1662: II.16). But there are impure definitions as well. *Lexical* definitions are intended to capture the ordinary meaning of the word and are consequently discovered and not stipulated by lexicographers (1662: I.14). As lexical definitions are atypical nominal definitions, *descriptions* are atypical real definitions. Descriptions list a number of accidents which are sufficient to specify the extension of an idea, but not the common nature of things within the extension (1662: II.16).

According to Arnauld and Lancelot there are two major categories of words: those that signify the objects of thoughts and those that signify the manner of thoughts (1660: II.1, II.13). The most important subcategory within the first is that of nouns; within the second, verbs. Concrete substantival nouns (such as 'sun') signify substances, abstract substantival nouns (such as 'whiteness') signify modes (1662: II.1). An adjectival noun has two significata: 'white' signifies distinctly the mode of whiteness, and confusedly signifies white things. A verb is a word whose principal function is to indicate assertion, and the copulative verb 'to be' has only this principal function (1660: II.13). Other verbs also have a secondary role: they express ideas and thereby refer to entities, and they indicate the time with respect to which the assertion is made. For instance, the verb 'lives' in the sentence 'Peter lives' expresses the mode of being alive and indicates the assertion that this mode belongs to Peter at the time of the assertion (1660: II.13).

The mind judges by uniting or separating two ideas. The product of judging is a proposition. Propositions are *simple* if they have a single subject and a single predicate, and *compound* otherwise. Compound propositions include among others conjunctions, disjunctions and conditionals. Not all complexity in the subject or predicate terms indicates that the proposition is compound. For example, 'The invisible God has created the visible world' - or equivalently, 'God who is invisible created the world which is visible' - is a *simple but complex* proposition. For although it contains three propositions - 'God is invisible', 'God created the world' and 'The world is visible' - only the second of these is asserted, while the others are assumed or taken for granted ((1662: II.5; 1660: II.9). By contrast, the proposition 'God is invisible and he created the visible world' is compound, since it ascribes two different predicates to the same subject (see Arnauld, A.; Port-Royal).

4 Locke's *Essay*

Although remaining firmly in the empiricist tradition, John Locke's *An Essay Concerning Human Understanding* (1689) incorporates some of the insights of Cartesian rationalism. This synthesis makes the *Essay* the essential point of reference for all philosophy of language in the eighteenth century.

According to Locke, the primary purpose of language is communication. God or nature made our organs fit to produce articulate sounds, and through arbitrary imposition we are capable of making these sounds 'Signs of internal Conceptions' (Locke 1689: III.i.2; see also III.ii.1). At first glance, this view resembles Hobbes', but - as

Leibniz remarks at the beginning of Book III of his *Nouveaux essais sur l'entendement humain (New Essays on Human Understanding)* (posthumously published in 1765) - there is an important difference: for Hobbes the private use of words is fundamental, whereas for Locke it is their use as signs for others.

The main thesis of Locke's theory of signification is that 'Words in their primary or immediate signification, stand for nothing, but the Ideas in the Mind of him that uses them' (1689: III.ii.2). Since the purpose of words is to invoke in the hearer's mind an idea identical (or at least sufficiently similar) to the idea the speaker has in mind, a word is bound to be intimately connected to an idea of the speaker. To say that primary signification is nothing but the speaker's idea does not mean that the word refers to that idea, or that in using a certain word the speaker is talking about the idea his word primarily signifies. Besides the primary signification of words, men often 'in their Thoughts give them a secret reference to two other things': the ideas in the minds of other speakers, and the reality of things (1689: III.ii.4). This supposition of *secondary signification* is indispensable for human beings, since 'without this double Conformity of their Ideas, they find, they should both think amiss of Things in themselves, and talk of them unintelligibly to others' (1689: II.xxxii.8). Locke apparently condemns only the assumption that words can directly signify the ideas of others and the reality of extramental things, without the mediation of their primary signification.

In accordance with the Cartesian tradition, Locke emphasizes that a genuine language cannot contain only singular names. We need signs which stand for more than one thing, since without these there would be too many (perhaps infinitely many) words to learn 1689: III.i.3; III.iii.2-4). With regard to the signification of these general terms Locke was a conventionalist. His primary target is the doctrine that general terms signify substantial forms, genuine universals in the things themselves that provide the ground for an objective classification of things. Since, he argues, our ideas are made up exclusively from the materials provided by experience (1689: II.xii.1), and since we cannot form ideas of substantial forms in this manner (1689: III.vi.10), substantial forms cannot be the significata of general words. The general term 'horse' signifies the *abstract general idea* of a horse, an idea that represents actual horses in virtue of representing certain sensible qualities they have in common. According to Locke, when we categorize things, we use the significata of our general terms. Fixing the significata of these terms is not a matter of discovering determinate species that exist independently of us, but rather a matter of reasonable choice from among the innumerable objective similarities of things (1689: III.iii.13).

In the case of ideas of substance, which represent natural unions of qualities, our forming of the appropriate abstract idea is independent of our linguistic abilities. The case is more complicated in the case of ideas of mixed modes, which represent collections of qualities that are not necessarily united in the same subject. For such complex ideas, although it is 'the Mind that makes the Collection, 'tis the Name which is, as it were the Knot, that ties them fast together' (1689: III.v.10). In the case of simple modes, like those of number, the dependence on naming is even more explicit: according to Locke, we could not have ideas of large numbers for which we have no names (1689: II.xvi.5).

Locke subscribes to the Port-Royal view that syncategorematic words - terms like 'is' and 'but', which he calls *particles* - stand not for ideas but for operations of the mind by which the ideas signified by other words are put together into a proposition. The truth of a proposition consists in the 'joining or separating of Signs, as the Things signified by them, do agree or disagree with one another' (1689: IV.v.2). Whether for Locke truth consists always in agreement and disagreement between ideas, or whether he allows for agreement and disagreement between an idea and what it represents, is a controversial question of Locke interpretation (see LOCKE, J. §§3-5).

5 Critique of Locke: Berkeley and Leibniz

The main tenets of Lockean philosophy of language were widely accepted among philosophers in the eighteenth century, but they did not remain unchallenged. The most important objections to the Lockean theory were made by George Berkeley and Gottfried Wilhelm Leibniz. Berkeley's critique was sketched in his *Philosophical Commentaries* (written between 1707 and 1708), and elaborated in *A Treatise Concerning the Principles of Human Knowledge* (1710), and his dialogue *Alciphron* (1732). Leibniz's objections take the form of a line-by-line critical commentary in his *New Essays on Human Understanding* (1765).

Berkeley's critique focuses on three major areas: the purpose of language, the semantics of general terms, and the nature of signification. He denies that communication is the chief end of discourse. Besides conveying our

thoughts to others, language is also used in 'the raising of some passion, the exciting to, or deterring from an action, the putting the mind in some particular disposition' (Berkeley 1710: Introduction 20). These are direct effects of speech, possibly achieved without the mediation of ideas. Berkeley's remark concerning the use of the expression 'good thing' prefigures later emotivist treatments: the function of such expressions is to excite the appropriate feelings in us (1710: Introduction 20) (see Emotivism). Even expressions that - unlike 'good thing' - do signify ideas, can do so in the way the letters of algebra signify, where 'though a particular quantity be marked by each letter, yet to proceed right it is not requisite that in every step each letter suggest to your thoughts, that particular quantity it was appointed to stand for' (1710: Introduction 19).

Berkeley harshly criticizes a view he associates with Locke concerning the generality of words and ideas. For Berkeley, a general word stands for a class of particulars by signifying 'an idea, which considered in itself is particular, [but] becomes general, by being made to represent or stand for all other particular ideas of the same sort' (1710: Introduction 12). There is no need to assume that a single abstract idea is signified by each general term: by the mechanism of *selective attention* we can consider one or another feature of a particular idea, and then use it as a sign for all particular ideas that share the given feature. Berkeley's criticism may miss its target. When Locke talks about abstraction, he also uses the language of considering certain features within a particular idea (for example, (Locke 1689: II.xi.9). Still, it remains unclear whether Locke thinks that abstract ideas are made from particulars via selective attention, or whether he in fact agrees with Berkeley that abstract ideas simply are particular ideas considered selectively.

Berkeley rejects the thesis that certain words have a secondary signification without the mind; things without the mind could be either material or immaterial, and Berkeley denies the existence of the former and the imaginability of the latter. Words like 'horse' or 'white' signify an idea without indicating anything beyond the idea; words like 'soul' or 'God' signify a substance without the mediation of an idea. A language is a network of signification relations between ideas which obtain in virtue of some customary connection. If the connections are the result of human imposition (for example, associating the idea of the letter 'a' with the idea of the corresponding sound, or associating the idea of the word 'apple' with some idea of an apple), the system of significations is an *artificial language*. If the connections result from divine imposition (for example, the connection between the ideas of fire and smoke, or between the visual idea of a square and a tactile idea of a square), the system is a *natural language* (Berkeley 1733: 40; see also Berkeley 1709: 140, 147). For Berkeley, the study of nature is the study of grammar for the divine language. Natural phenomena - which, of course, are merely collections of ideas - can signify one another, but they cannot stand in genuine causal relations to one another: it is God's creating and sustaining act that causes everything to exist and be the way it is (1733: 13; 1710: 65, 108) (see Berkeley, G. §§5-6, 11).

Leibniz's two main targets in the Lockean theory are claims wholeheartedly endorsed by Berkeley: that signification of words is arbitrary and that categorization is the work of the mind. According to Leibniz, signification is 'settled by reasons - sometimes natural ones in which chance plays some part, sometimes moral ones which involve choice' (Leibniz 1765: III.2). To support this thesis, Leibniz presents etymological links connecting contemporary words with earlier forms, and sometimes ultimately with onomatopoeia. What grounds these explanations is the principle of sufficient reason: brute accidents are as unacceptable in semantics as in physics.

Leibniz defends the semantic relevance of real definitions against Locke. Although 'gold' can be given a number of different nominal definitions, only a few of these will be in agreement with the genuine conceptual analysis. A real definition is grounded in a real species, and has the characteristic mark that it 'displays the possibility of the definiendum' (1765: III.3). Since he rejects the view that ideas we have must be completely transparent to us, he can resist Locke's claim that our words cannot signify unknown features: 'the name "gold"… signifies not merely what the speaker knows of gold, e.g. something yellow and very heavy, but also what he does not know, which may be known about gold by someone else, namely: a body endowed with an inner constitution from which flow its color and weight, and which also generates other properties which he acknowledges to be better known by the experts' (1765: III.11). If pressed, we would have to come up with nominal definitions for substances; but these do not fix the meaning of the corresponding terms, since we recognize that 'our definitions are all merely provisional'. We know that something that passes all our current tests for being gold may nevertheless fail to be gold, since 'one might… discover a new assaying method which would provide a way of distinguishing natural gold from this artificial gold' (1765: III.4, 6).

What underlies this insistence on real definitions is Leibniz's commitment to the ideal of a *universal characteristic*, a written language which could be understood by any rational being and which could perfect human reasoning itself. Corresponding to each complex concept or proposition, there is a complex symbol in the universal characteristic whose structure is the same as the structure of the concept or proposition signified. Occasionally, Leibniz talks about the primitive symbols of the universal characteristic as geometrical figures - like Egyptian hieroglyphs or Chinese characters - which bear some similarity to what they represent Liebniz 1875-90: IV.72-3). He also made attempts to assign numerical values to various concepts and to represent conceptual analysis as analogous to prime factorization (Couturat 1903: 245-6). But - as Leibniz himself realized - our concepts are so complex that we have no reason to believe that the process of resolving them into primitive constituents is a finite one. This does not mean that the system of numerical characteristic is useless. Since logic imposes constraints on possible assignments of numerical values to concepts, and since these constraints are expressible in the form of Diophantine equations, the system of numerical characteristic without the analysis of concepts can be used as a *formal calculus* to check the validity of inferences (1875-90: VII.205) (see Leibniz, G.W. §10).

6 The origin of language: Condillac and Rousseau

Questions about the origin of language were widely discussed in the seventeenth century but the views proposed did not go much beyond the biblical or the Lucretian accounts. This changed radically in 1746 with the publication of Étienne Bonnot de Condillac's *Essai sur l'origine des connaissances humaines (An Essay on the Origin of Human Knowledge)*.

Condillac was an enthusiastic but critical follower of Locke. His disagreement with Locke concerns two issues related to language. Condillac's first criticism is that Locke's theory about the mind is merely descriptive and fails to explain how understanding originates from sensation. The second is that Locke underestimates the role language plays in a theory of ideas. The two points are connected, since the 'the consideration of words, and of our manner of using them, might give some light into the principle of our ideas' and could yield 'a far better explication of the springs of the human understanding' (introduction to Condillac 1746).

Condillac claims that human beings can attain control over their thoughts solely in virtue of the use of *instituted signs*. Animals lack the ability to establish a conventional relation between an object and its sign, they cannot recall ideas at will, and therefore cannot prudently direct their behaviour. Animals are driven by instinct, which is 'no more than the imagination, as independent of our command, though by its activity, completely concurring to the preservation of our being' (1746: I.ii.95). In *Traité des sensations (Treatise on Sensation)* (1754), Condillac imagines a statue gradually brought to life by obtaining its five senses one by one. According to Condillac, the fully sensitized statue will be able to recognize its basic needs, remember things and learn from experience but, because it lacks speech, it will have only 'the habit of directing itself by ideas for which it does not know how to account' (1754: II.viii.35).

But it seems that in order to make use of instituted signs, we must already have sufficient reflection to choose them and to establish the conventional link between the signs and their significata (1746: I.ii.49). This is the problem that Condillac's account of the origin of language is supposed to answer. The explanation begins with natural signs - cries and gestures: these constitute the *language of action* (1746: II.i.1). The development of reflection and the development of language are intertwined: since both require the other, they evolve through a process of bootstrapping. Communicative intentions first arise due to the recognition of the accidental success of the language of action. From that point on, reflection and language interact to the mutual expansion and sophistication of both (see Condillac, É. B. de §§2, 3).

In the years following the appearance of Condillac's *Essay*, Diderot, Maupertuis and Turgot in France, Adam Smith and Lord Monboddo in England, Moses Mendelssohn and Johann Gottfried Herder in Germany each expressed their views on the origin of language. But perhaps the most significant contributions to the debate initiated by Condillac were Jean-Jacques Rousseau's *Discours sur l'origine et les fondements de l'inégalité parmi les hommes (Discourse on the Origin and Foundations of Inequality among Mankind)* (1755) and *Essai sur l'origine des langues (On the Origin of Languages)* (published posthumously in 1781). Rousseau replaces Condillac's relatively harmonious conception of the course of development with his emphasis on the conflict between human nature and social existence. He agrees that languages arise because of human needs, but he rejects the idea that these are physical needs. Physical needs would never have forced us to abandon the language of

gesture (Rousseau 1781: 1). What stimulated the first words were *moral needs*, passions. 'As man's first motives for speaking were of the passions, his first expressions were tropes' (1781: 3). That is, language and art were not separated at the beginning of history. As savage man turned into barbarian man and finally into civilized man, human needs changed, and as a consequence human languages lost their connection to music, becoming more and more rational and less and less passionate (1781: 19). And as societies assumed their final form, where 'no longer is anything changed except by arms and cash', languages have also become less favourable to liberty (1781: 20) (see Rousseau, J.-J. §2).

See also: Language, philosophy of; Language, Renaissance philosophy of; Universal language

ZOLTAŃ GENDLER SZABÓ

List of works

Arnauld, A. and Lancelot, C. (1660) *Grammaire générale et raisonnée de Port-Royal*, repr. Geneva: Slatkine, 1968; trans. J. Rieux and B.E. Rollin, *General and Rational Grammar: The Port-Royal Grammar*, The Hague: Mouton, 1975.(The Port-Royal account of the principles of grammar.)

Arnauld, A. and Nicole, P. (1662) *La logique ou l'art de penser*, ed. P. Clair and F. Girbal, Paris: Vrin, 1981; trans. J.V. Buroker, *Logic or the Art of Thinking*, Cambridge: Cambridge University Press, 1996. (The Port-Royal account of the principles of logic.)

Berkeley, G. (1707-8) *Philosophical Commentaries*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Berkeley's previously unpublished early notebooks.)

Berkeley, G. (1709) *An Essay towards a New Theory of Vision*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Berkeley's first published work: its subject matter is visual perception.)

Berkeley, G. (1710) *A Treatise Concerning the Principles of Human Knowledge*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Berkeley's main philosophical work.)

Berkeley, G. (1732) *Alciphron, or the Minute Philosopher*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Dialogue of philosophical, moral and theological content.)

Berkeley, G. (1733) *The Theory of Vision, Vindicated and Explained*, in A.A. Luce and T.E. Jessop (eds) *The Works of George Berkeley, Bishop of Cloyne*, London: Nelson & Sons, 1948-57, 9 vols.(Reply to criticism of Berkeley 1709.)

Condillac, É.B. de (1746) *Essai sur l'origine des connaissances humaines*, in G. Le Roy (ed.) *Oeuvres philosophiques*, Paris: Presses Universitaires de France, 1947-51, 3 vols; trans. T. Nugent, *An Essay on the Origin of Human Knowledge*, Gainesville, FA: Scholar's Facsimile and Reprints, 1971.(Condillac's main philosophical work.)

Condillac, É.B. de (1754) *Traité des sensations*, in G. Le Roy (ed.) *Oeuvres philosophiques*, Paris: Presses Universitaires de France, 1947-51, 3 vols; trans. G. Carr, *Treatise on Sensations*, Los Angeles: University of Southern California Press, 1930.(A treatise on the nature of sense perception.)

Descartes, R. (1637) *Discours de la méthode*, in C. Adam and P. Tannery (eds) *Oeuvres de Descartes*, Paris: Vrin/C.N.R.S., 1964-76, 11 vols.(The standard edition of Descartes' writings.)

Hobbes, T. (1651) *Leviathan*, ed. R. Tuck, Cambridge: Cambridge University Press, 1991.(Hobbes' main work in political philosophy.)

Hobbes, T. (1655) *De corpore*, in W. Molesworth (ed.) *The Latin works of Thomas Hobbes of Malmesbury*, London: John Bohn, 1845, 5 vols; trans. W. Molesworth, *The English Works of Thomas Hobbes of Malmesbury*, London: John Bohn, 1839-45, 11 vols. (Hobbes' main work in natural philosophy.)

Leibniz, G.W. (1765) *Nouveaux essais sur l'entendement humain*, in German Academy of Sciences (ed.) *G.W. Leibniz: Sämtliche Schriften und Briefe*, Berlin: Akademie, 1923-; trans. P. Remnant and J. Bennett, *New Essays on Human Understanding*, Cambridge: Cambridge University Press, 1981.(A paragraph-by-paragraph commentary and critical discussion of Locke 1689.)

Leibniz, G.W.(1875-90) *Die Philosophische Schriften von Leibniz*, ed. C.I. Gerhardt, Berlin: Weidmann, 7 vols. (One of the standard editions of Leibniz's writings.)

Leibniz, G.W. (1903) *Opuscules et fragments inédits de Leibniz*, ed. L. Couturat, Paris: Alcan.(A collection of smaller, previously unpublished pieces by Leibniz.)

Locke, J. (1689) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press, 1975.

(Locke's main philosophical work.)

Locke, J. (1690) *Two Treatises of Government*, ed. P. Laslett, Cambridge: Cambridge University Press, 1960. (Locke's main work in political theory.)

Lucretius (*c.*55 BC) *On the Nature of Things*, ed. C. Bailey, *Titi Lucreti Cari De rerum natura libri sex*, Oxford: Oxford University Press, 3 vols, 1947.(Bailey is the monumental standard edition, with text, translation and commentary, still highly informative and not fully superseded.)

Rousseau, J.-J. (1755) *Discours sur l'origine et les fondements de l'inégalité parmi les hommes (Discourse on the Origin and Foundations of Inequality among Mankind)*, in B. Gagnebin and M. Raymond (eds) *Oeuvres Complètes de Jean-Jacques Rousseau*, vol. 3, Paris: Pléiade, 1964; trans. D.A. Cress, *Jean-Jacques Rousseau: Basic Political Writings*, Indianapolis, IN: Hackett Publishing Company, 1987.(Rousseau's prize-winning essay on the origins of inequality.)

Rousseau, J.-J. (1762) *Du contrat social (The Social Contract)*, in B. Gagnebin and M. Raymond (eds) *Oeuvres Complètes de Jean-Jacques Rousseau*, vol. 3, Paris: Pléiade, 1964; trans. D.A. Cress, *Jean-Jacques Rousseau: Basic Political Writings*, Indianapolis, IN: Hackett Publishing Company, 1987.(Rousseau's most famous political writing.)

Rousseau, J.-J. (1781) *Essai sur l'origine des langues*, ed. C. Porset, Bordeaux: Ducros, 1968; trans. in J.H. Moran and A. Gode (eds) *On the Origin of Languages*, Chicago, IL: University of Chicago Press, 1966. (Rousseau's main contribution to the debate about the origin of language in the eighteenth century.)

References and further reading

Aarsleff, H. (1982) *From Locke to Saussure*, Minneapolis, MN: University of Minnesota Press.(Collection of essays on Wilkins, Locke, Leibniz, Condillac, Herder and lesser figures.)

Bertman, M. (1978) 'Hobbes on Language and Reality', *Revue Internationale de Philosophie* 32: 536-51.

Buroker, J.V. (1993) 'The Port-Royal Semantics of Terms', *Synthèse* 96: 455-75.

Chomsky, N. (1966) *Cartesian Linguistics*, New York: Lanham.(Controversial account about the connections between modern generative grammar and the Port Royal.)

Ducret, M. and Lannay, M. (1967) 'Synchronie et diachronie: *L'essai sur l'origine des langues* et le second *Discours*', *Revue Internationale de Philosophie* 21: 421-42.(On Rousseau's views about the origin of language. The entire issue of the journal is devoted to philosophy of language in the eighteenth century.)

Flage, D. (1987) *Berkeley's Doctrine of Notions: A Reconstruction Based on his Theory of Meaning*, London: Croom Helm.

Guyer, P. (1994) 'Locke's Philosophy of Language', in V. Chappell (ed.) *The Cambridge Companion to Locke*, Cambridge: Cambridge University Press.(Survey article. The volume contains a detailed bibliography.)

Heinekamp, A. (1975) 'Natürliche Sprache und Allgemeine Charakteristik bei Leibniz', *Studia Leibnitian Supplementa* 15: 257-86. (Looks at natural language and the *universal characteristic* in Liebniz.)

Hungerland, I. and Vick, G.R. (1973) 'Hobbes's Theory of Signification', *Journal of History of Philosophy* 11: 459-82.

Ishiguro, H. (1972) *Leibniz's Philosophy of Logic and Language*, Cambridge: Cambridge University Press. (Includes discussion of the connections between Leibniz's views on language and his metaphysics.)

Juliard, P. (1970) *Philosophies of Language in Eighteenth Century France*, The Hague: Mouton.(Discusses the philosophy of language of Condillac, Maupertuis, La Mettrie, Turgot and Rousseau.)

Kremer, E.J. (ed.) (1994) *The Great Arnauld and Some of His Philosophical Correspondents*, Toronto, Ont.: University of Toronto Press.(A collection of papers with detailed bibliography.)

Kretzmann, N. (1968) 'The Main Thesis of Locke's Semantic Theory', *Philosophical Review* 77: 175-96.

Land, S. (1986) *The Philosophy of Language in Britain*, New York: AMS Press.(Discussion of the philosophy of language of Hobbes, Locke, Berkeley, Harris, Reid, Monboddo and Smith.)

Nuchelmans, G. (1983) *Judgment and Proposition from Descartes to Kant*, Amsterdam: North Holland.

Porset, C. (1977) 'Grammatista philosophans. Les sciences du langage de Port-Royal aux Idéologues (1660-1818)', in A. Joly and J. Stefanini (eds) *La Grammaire générale. Des Modistes aux Idéologues*, Lille: Presses Universitaires de France, 11-95.(Bibliography of works in the philosophy of language and grammar between 1660 and 1818. Also covers the secondary literature before 1976.)

Ricken, U. (1984) *Linguistics, Anthropology and Philosophy in the French Enlightenment*, London: Routledge. (General survey, includes a detailed bibliography.)

Rousseau, N. (1986) *Connaissance et langage chez Condillac*, Geneva: Libraire Droz.(Monograph on Condillac with a detailed bibliography.)

Rutherford, D. (1995) 'Philosophy and Language in Leibniz', in N. Jolley (ed.) *The Cambridge Companion to*

Leibniz, Cambridge: Cambridge University Press, 224-69.(Survey article. The volume contains a detailed bibliography.)

Sgard, J. (ed.) (1982) *Condillac et les problèmes du langage*, Geneva: Slatkine.(Collection of critical essays.)

Slaughter, M. (1982) *Universal Languages and Scientific Taxonomy in the Seventeenth Century*, Cambridge: Cambridge University Press.

Winkler, K.P. (1989) *Berkeley: An Introduction*, Oxford: Clarendon Press.(Chapters 1 and 2 deal with questions of language and abstraction in Berkeley.)

Language, Indian theories of

Language is a much debated topic in Indian philosophy. There is a clear concern with it in the Vedic texts, where efforts are made to describe links between earthly and divine reality in terms of etymological links between words. The earliest surviving Sanskrit grammar, Pāṇini's intricate Aṣṭādhyāyī(Eight Chapters), dates from about 350 BC, although arguably the first explicitly philosophical reflections on language that have survived are found in Patañjali's 'Great Commentary' on Pāṇini's work, the Mahābhāṣya (c.150 BC). Both these thinkers predate the classical systems of Indian philosophy. This is not true of the great fifth-century grammarian Bhartṛhari, however, who in his Vākyapadīya (Treatise on Sentences and Words) draws on these systems in developing his theory of the sphoṭa, a linguistic entity distinct from a word's sounds that Bhartṛhari takes to convey its meaning.

Among the issues debated by these philosophers (although not exclusively by them, and not exclusively with reference to Sanskrit) were what can be described as (i) the search for minimal meaningful units, and (ii) the ontological status of composite linguistic units. With some approximation, the first of these two issues attracted more attention during the early period of linguistic reflection, whereas the subsequent period emphasized the second one.

1 Historical sketch

Linguistic science in India started soon after the Vedic period. The earliest grammarian whose work has survived is Pāṇini (*c*.350 BC), author of the *Aṣṭādhyāyī(Eight Chapters)*. This work consists of some 4,000 aphoristic statements (*sūtras*) which describe the Sanskrit language in considerable detail, but leave no space for explicit reflections about the nature of language. Such reflections make their appearance in the voluminous *Mahābhāṣya (Great Commentary)* of Patañjali (*c*.150 BC). The *Mahābhāṣya* is a commentary on the *Aṣṭādhyāyī* (but not on all of its *sūtras*), and on the aphoristic *vārttikas* of Kātyāyana, which comment upon Pāṇini's *sūtras*. Another linguist whose work has been preserved and who, like Kātyāyana, appears to belong to the period between Pāṇini and Patañjali, is Yāska, author of the *Nirukta (Etymological Explanation)*.

All these authors precede the formation of the classical systems of Indian philosophy; their reflections on language are therefore largely unaffected by them. This changes with Bhartṛhari (*c.* fifth century), perhaps the first commentator on Patañjali's *Mahābhāṣya*, and the author of the *Vākyapadīya (Treatise on Sentences and Words)*. Bhartṛhari is well aware of the philosophies of his time, and makes ample use of them to construct his own system, which he presents as the philosophy of grammar. The subsequent Pāṇinian tradition accepts this philosophy (or what it preserves of it) as its own, but there are remarkably few grammarians who write treatises on it. Apart from the three principal commentators on the *Vākyapadīya* - Helārāja (tenth century), Puṇyarāja and Vṛṣabhadeva (dates unknown) - by far the most important among them are Kauṇḍa Bhaṭṭa and Nāgeśa Bhaṭṭa, both belonging to the most recent period of grammatical studies (after 1600). Some authors belonging to different schools of thought, however, adopt and defend some of the points of view of the grammarians. The ontological status of composite linguistic units is a subject that evokes special interest.

2 The search for minimal meaningful units

The different linguistic sciences of ancient India - and in particular grammar and etymological explanation - have to be understood against the background of the practice, common in the Vedic Brāhmaṇas (before Pāṇini), of giving etymological explanations of names of gods and of other terms, usually related to the sacrifice and often occurring in sacred formulas (*mantras*). Unlike those of modern linguistics, the etymologies of ancient India have nothing to do with the origin or the history of the words concerned. They cannot, because language, and the Sanskrit language in particular, was looked upon as stable in time; from the subsequent period we know that many even believed Sanskrit to be eternal, that is, without beginning. These etymologies establish links between things and the mythological reality that hides behind them. The god Agni ('fire'), for example, is thus called because he was created first (*agre*). There are countless etymologies of this kind in the Brāhmaṇas. These texts frequently add that the gods have obscured a number of the etymological links. The god Indra, for example, is 'really' called Indha ('the kindler') because he kindled the vital airs. However, people call him Indra because the gods are fond of the cryptic, and dislike the evident. Knowledge of the links revealed by etymologies is important in reaching one's religious goals.

Yāska's *Nirukta* takes the validity of such etymologies for granted, but secularizes their use. It presents etymologizing as a way to arrive at the meaning of unknown words. Moreover, only nouns and adjectives can be etymologically explained, and then only in terms of verbal forms: verbs explain nominal words, and show the (or an) activity that characterizes the object named. Yāska illustrates his method with the help of known words. One might expect that this procedure would lead to the identification and isolation of the common parts found in different words (such as the common part *ag* of *agni* and *agre* in the above example), and would determine their meanings. But Yāska's demands with regard to the semantic adequacy of etymological explanations are so stringent that this turns out to be impossible. He insists, for example, on two different etymological explanations for words that have two meanings. This rigour forces Yāska to be very undemanding with respect to the phonetic similarity required in etymological explanations. With the help of a number of examples from grammar, he shows that phonemes may disappear, be modified, change position, and so on. The same applies, *a fortiori*, to etymology. Similarities between words in etymological explanations may, as a consequence, be minimal: one single phoneme in common may have to do. The main thing is that one should not be discouraged; one should not stop looking for an etymological explanation simply because one does not find similar words.

It is interesting to observe that Patañjali's *Mahābhāṣya* contains a passage which, like Yāska's *Nirukta*, shows how phonemes may undergo change of position, elision and modification in grammatical derivations. Unlike Yāska, Patañjali concludes from this quite explicitly that phonemes by themselves cannot have meaning, although it seems likely that Yāska, even though he does not state it in his *Nirukta*, drew the same conclusion as Patañjali. It appears that Yāska's semantic rigour prevented him from trying to identify the ultimate meaningful constituents of Sanskrit (as we see for ancient Greek in Plato's *Cratylus*, for example). This task, but on a far less ambitious scale, is left to the grammarians, among whom Pāṇini is the most famous. His grammar is not an analysis of Sanskrit but a synthesis: it produces the words and sentences of the language, starting from their ultimate meaning-bearing constituents, essentially stems and affixes. To be precise, Pāṇini's grammar first furnishes stems and affixes on the basis of a semantic input, and these stems and affixes are subsequently joined together, and modified where necessary, so as to yield words and sentences (see Bronkhorst 1980).

In view of the background of Vedic etymologies, discussed above, one may legitimately conclude that Pāṇini considers these ultimate meaningful constituents to be really meaningful, more so perhaps than the 'surface forms' they help produce. The search for 'really meaningful' ultimate constituents of language is clearly present in the efforts of the grammarians. Pāṇini's *sūtra* 1.2.45, which recognizes but three meaningful entities, namely verbal roots, nominal stems and suffixes, indicates that words and sentences are considered to have at best a composite meaning. This search manifests itself later in the attribution by a number of Tantric thinkers of metaphysical meanings to individual phonemes. They can afford to go further than Pāṇini and Yāska in their analysis, continuing all the way to the individual phonemes, owing to the fact that they are less limited by semantic considerations. This in its turn is no doubt linked to the circumstance that sacred formulas (*mantras*) in Tantrism (unlike the Vedic ones) have shed their connection with ordinary language and its semantic constraints.

Returning to Pāṇini and grammatical analysis, later grammarians, mainly under the influence of Patañjali and Bhartṛhari, reject the position according to which the ultimate meaningful constituents presented by grammar are somehow more real than the words they produce. For them, stems and affixes are conventions, or rather inventions of grammarians. This reaction is to be understood in the light of the ontological concerns to be discussed below. The semantic analysis underlying Pāṇini's procedure, on the other hand, came to be generally accepted (albeit sometimes with slight modifications). Later thinkers use this analysis as the basis for deliberations on the relative importance of the various 'semantic elements' that Pāṇini assigns to a sentence in the understanding obtained by a hearer (*śābdabodha*). In a sentence like *caitraḥ pacati* ('Caitra cooks'), to take a simple example, the grammatical elements are: *caitra-s pac-a-ti*. Of these, the following are expressive: *caitra*, *pac* and *ti*. Thinkers of the new Nyāya school (Navya-Nyāya) consider the grammatical subject (in this case *caitra*) most important, and give (approximately) the following semantic analysis of the sentence: 'Caitra characterized by the activity of cooking'. The grammarians look upon the meaning of the verbal root (*pac*) as central, and paraphrase the sentence (again approximately) as: 'The activity of cooking whose agent is Caitra'. The Mīmāṃsakas, finally, put emphasis on the verbal suffix (here *ti*); since they are primarily interested in Vedic injunctions, and consequently in imperative and optative verbal forms, we shall not enter into the details of their analyses.

3 The ontological status of composite linguistic units

To appreciate the importance of the debate on the ontological status of composite linguistic units, one has to be aware of the great interest in ontological questions that characterizes much of Indian philosophy. In the realm of language this leads to questions like: Do words and sentences really exist? If so, how can they, given that the phonemes that constitute them do not occur simultaneously? Since, moreover, simultaneous occurrence is a condition for the existence of collective entities, do individual phonemes exist? They, too, have a certain duration, and consist therefore of parts that do not occur simultaneously.

Perhaps the first to address these questions were Buddhists of the Sarvāstivāda school. These Buddhists were active in the first centuries BC in drawing up lists of elements - the so-called *dharmas* - which were considered to constitute all there is (see Buddhism, Ābhidharmika schools of). The list accepted by the Sarvāstivādins contains three elements which correspond to phonemes (*vyañjanakāyas*), words (*nāmakāya*) and sentences (*padakāya*) respectively. This means that these Buddhists postulated phonemes, words and sentences as existing entities which, like virtually everything else in their ontology, are momentary. Little is known about the way they visualized the mutual relationship between these entities, or how they would answer the questions formulated above.

The grammarian Patañjali may have been influenced by these ideas. He certainly knew the notion of an individual phoneme and of a word conceived of as a single entity. For Patañjali, these phonemes and words are not momentary; they are, on the contrary, eternal. One should not, however, attach too much importance to this difference: for the Buddhists, everything is momentary; for many Brahmans, the Veda, and therefore also its language, is eternal. It is more important to observe that these notions play a relatively minor role in Patañjali's expositions. They acquire major significance in Bhartṛhari's *Vākyapadīya*, where they are made to fit his general philosophy that more comprehensive totalities are more real than their constituent parts.

It appears that in the period between Patañjali and Bhartṛhari a major shift of emphasis took place in the discussion of linguistic units. The discussion became centred on the linguistic unit as meaning-bearer. The problem of individual phonemes, which have no meaning, came to be separated from that of words, grammatical elements (stems and affixes) and sentences, which do. In the context of Bhartṛhari's philosophy this is understandable, for here linguistic units and the 'objects' they refer to are treated in a parallel fashion. But this shift of emphasis was not confined to the grammatical tradition. Śabarasvāmin, the author of the oldest and most important surviving commentary on the *Mīmāṃsāsūtra*, and who may be an approximate contemporary of Bhartṛhari, cites (1.1.5; see Frauwallner 1968: 38-) an earlier commentator who rejects the notion of a word as different from its constituent phonemes. This does not, however, prevent him from proclaiming that phonemes are single and eternal. In other words, phonemes and words undergo a different treatment altogether. Moreover, the author of the *Yogabhāṣya* (whose name was probably Patañjali, like the author of the *Mahābhāṣya*, although he was certainly different from the latter; recent tradition calls him Vyāsa) speaks about the single word which is without parts, without sequence, without constituent phonemes, and which is mental (on *Yogasūtra* 3.17). This Patañjali may have lived around 400 AD, and therefore perhaps before Bhartṛhari.

4 Early *sphoṭa* theory

Patañjali (the grammarian) and Bhartṛhari use the word *sphoṭa* to refer to linguistic entities conceived of as different from the sounds that reveal them. For Patañjali, the *sphoṭa* does not necessarily convey meaning; he uses the term also in connection with individual phonemes. For Bhartṛhari, the *sphoṭa* is a meaning-bearer. The *sphoṭa*, he points out, is different from the sounds which manifest it, and he makes several suggestions as to what constitutes it. It might be a mental entity. Or one might take it to be the universal residing in the manifesting sounds. One could even look upon the material basis of words, for example, wind, as being the *sphoṭa*. Bhartṛhari presents these options, but his perspectivism allows him to avoid choosing between them.

Arguments claiming to prove the existence of the *sphoṭa*, as well as arguments which try to refute it, henceforth concentrate heavily, even exclusively, on the *sphoṭa* as meaning-bearing unit. The primary question is not 'What exactly is the *sphoṭa*?' but rather 'How can a sequence of phonemes, each without meaning and not even occurring simultaneously, express meaning?' According to some, a sequence of sounds can express meaning; they have to show how it does so. Others hold that this is not possible; they solve the problem by postulating the existence of

the *sphoṭa*. These two positions find their classical expositions in Kumārila Bhaṭṭa's critique of the *sphoṭa* doctrine in his *Ślokavārttika (Commentary in Verse)* (seventh century), and in Maṇḍana Miśra's (*c*.700?) defence against these attacks in his *Sphoṭasiddhi (Demonstration of the Sphoṭa)*. Neither Kumārila nor Maṇḍana were grammarians: the former belonged to the school of Vedic hermeneutics called Mīmāṃsā; the latter, too, had links with this school.

Kumārila, elaborating the opinions of Śabarasvāmin (see §3), on whose *Mīmāṃsābhāṣya* he comments, accepts the eternal existence of individual phonemes. But he combats the notion that more than phonemes are required to understand the meaning-bearing function of language. It is true that the phonemes constituting a word are not pronounced simultaneously. But there are situations where everyone agrees that a series of activities that succeed each other in time can none the less jointly produce an effect. He gives the example of a Vedic sacrifice, whose constituent activities are performed at different times, but which produces a single result, namely heaven. Another example concerns counting: we can count objects in sequence, one after the other, and arrive at one result, their number. Furthermore, the fact that individual phonemes are without meaning does not exclude the possibility that they can express a meaning when pronounced in sequence. The parts of a cart, too, cannot fulfil the functions that a cart can fulfil. Last but not least, though the constituent phonemes of a word are not pronounced simultaneously, they are remembered together the moment the last phoneme is (or has just been) uttered.

Maṇḍana answers Kumārila's arguments one by one. He protests against the idea of the combined memory of the phonemes that constitute a word. First of all, one does not remember phonemes, but the word as a whole. Second, memory impressions can only present to us their contents, in this case phonemes, not something else, such as the meaning of the word. And third, two words may consist of the same phonemes, say 'pit' and 'tip' (a Sanskrit example is the pair *saraḥ*/*rasaḥ*, 'lake'/'taste'), so that the memories that combine their phonemes should be the same, yet they are recognized as different. Perhaps Maṇḍana's most interesting contribution to the discussion is his claim that the *sphoṭa* is directly perceived: it is gradually revealed by the phonemes.

The *sphoṭa* constitutes the central element of what came to be called the philosophy of the grammarians. All thinkers who deal with the issue, including Maṇḍana Miśra, refer in this connection to Bhartṛhari's *Vākyapadīya*. But the more encompassing ideas in the context of which Bhartṛhari worked out his ideas on the nature of linguistic entities largely escaped the attention of those who so faithfully cited him.

5 Later *sphoṭa* theory

After a lull, a revival of interest in the *sphoṭa* and related issues took place from the sixteenth century onward. Of the various authors who wrote treatises on the philosophy of grammar, Nāgeśa Bhaṭṭa (*c*.1700) was the most famous. He wrote a large number of treatises on various subjects, among them the *Sphoṭavāda (Exposition on Sphoṭa)*, the *Laghumañjūṣā(Small Casket)* and the *Paramalaghumañjūṣā(Extremely Small Casket)* (written in this order), which deal with the philosophy of grammar. These books show that Nāgeśa changed his mind several times with regard to the *sphoṭa* doctrine.

The *Sphoṭavāda* enumerates eight types of *sphoṭa*: (1) phoneme, (2) word, (3) sentence, (4) indivisible word, (5) indivisible sentence, (6) phoneme-universal, (7) word-universal, (8) sentence-universal. These *sphoṭas* are primarily meaning-bearers. The first and sixth ones, in spite of their misleading names, refer to grammatical elements (stems and affixes) rather than to phonemes. Nāgeśa's reasons for postulating these eight types are not always clear. This early work gives the impression that he collected various ideas without being able to combine them into one overarching vision.

This changes with the *Laghumañjūṣā*, which opens with the words: 'In this [work] the sentence-*sphoṭa* is most important.' Other parts of the work make it clear that Nāgeśa has been converted - no doubt under the influence of Bhartṛhari, whose *Vākyapadīya* he frequently cites - to the idea that only sentences really exist, that words and grammatical elements are no more than imaginary. He is particularly fierce with regard to grammatical stems and affixes.

Surprisingly, the *Paramalaghumañjūṣā*, meant to be an abbreviated version of the *Laghumañjūṣā*, begins, like the *Sphoṭavāda*, with an enumeration of the eight kinds of *sphoṭa*. Immediately following this it repeats the opening statement of the *Laghumañjūṣā*, according to which the sentence-*sphoṭa* is most important. Closer study reveals that Nāgeśa has been confronted with cases where the sentence-*sphoṭa* view comes into conflict with grammatical

derivations. There is a grammatical meta-rule which states that a grammatical derivation evolves in the order in which the expressive elements arise. Expressive elements acquire in this way importance, and it will not do to say that their expressiveness is merely imaginary. The issue is all the more important in view of the fact that there were different opinions as to whether the final substitutes of the grammatical elements - which appear in the 'surface forms' - are expressive, or whether the substituends are. This disagreement can have an effect on the correct derivation of words and sentences. Nāgeśa's final position is chosen in awareness of these complications. He still maintains that the sentence-*sphoṭa* is most important. But he no longer treats the other, 'imaginary', entities as lightly as he did earlier.

Nāgeśa is often thought of as the last great author in the Pāṇinian tradition. His vacillations where the *sphoṭa*-doctrine is concerned illustrate the conflict that exists between the two major issues of grammar distinguished in this entry: the search for minimal meaningful units on the one hand, and the ontological status of composite linguistic units on the other. His final position tries to give both their due: the idea inherited from Bhartṛhari that only the sentence is 'real', rather than words and smaller grammatical elements; and the idea inherited from Pāṇini that grammar is concerned with the smallest identifiable meaningful elements and the way they combine to form larger units.

See also: Interpretation, Indian theories of; Meaning, Indian theories of; MīmāṃsāČsect;3

JOHANNES BRONKHORST

References and further reading

Most of the ancient works listed here are highly technical and voluminous, and apt to be unrewarding for those who are not specialists in this field. Reliable translations are nonexistent in most cases.

Bhartṛhari (5th century) *Vākyapadīya (Treatise on Sentences and Words)*, ed. W. Rau, Wiesbaden: Franz Steiner, 1977.(A very difficult work that has occasioned a lot of controversy among scholars.)

Bronkhorst, J. (1980) 'The Role of Meanings in Pāṇini's Grammar', *Indian Linguistics* 40 (1979): 146-57.(Shows that meanings are the input of Pāṇini's grammar.)

Bronkhorst, J. (1981) '*Nirukta* and *Aṣṭādhyāyī*: Their Shared Presuppositions', *Indo-Iranian Journal* 23: 1-14. (Analyses the aims and procedures of Yāska's *Nirukta*.)

Bronkhorst, J. (1984) 'Patañjali and the Yoga Sūtras', *Studien zur Indologie und Iranistik* 10: 191-212.(§2 deals with the authorship of the *Yogabhāṣya*.)

Bronkhorst, J. (1986) *Tradition and Argument in Classical Indian Linguistics*, Dordrecht: Reidel.(On Nāgeśa Bhaṭṭa; see especially chapters 10 and 11, and appendix 3.)

Coward, H.G. and Raja, K.K. (eds) (1990) *Encyclopedia of Indian Philosophies*, vol. 5, *The Philosophy of the Grammarians*, Delhi: Motilal Banarsidass.(Contains bibliography up to 1983.)

Deshpande, M.M. (1992) *The Meaning of Nouns. Semantic Theory in Classical and Medieval India. Nāmārtha-nirṇaya of Kauṇḍabhaṭṭa*, Dordrecht: Kluwer.(Richly annotated translation of a chapter of Kauṇḍa Bhaṭṭa's *Vaiyākaraṇabhūṣaṇa*; the long introduction contains useful discussions of concepts of linguistic philosophy.)

Frauwallner, E. (1968) *Materialien zur ältesten Erkenntnislehre der Karmamīmāṃsā(Materials on the Oldest KarmamīmāṃsāEpistemology)*, Vienna: Hermann Böhlaus.(Contains text and translation of an important portion of Śabarasvāmin's *Mīmāṃsābhāṣya*.)

Gaurinath Sastri (1980) *A Study in the Dialectics of Sphoṭa*, Delhi: Motilal Banarsidass, revised edn.(Presents various arguments for and against the *sphoṭa* that were used in classical India.)

Houben, J. (1997) 'The Sanskrit Tradition', in W. van Bekkum *et al.*, *The Emergence of Semantics in Four Linguistic Traditions: Hebrew, Sanskrit, Greek, Arabic*, Amsterdam/Philadelphia, PA: John Benjamins, 49-145.(Survey of semantics in classical India.)

Kumārila Bhaṭṭa (7th century) *Ślokavārttika (Commentary in Verse)*, ed. Dvārikādāsa Śāstrī, Varanasi: Ratna Publications, 1978.(The most important commentary by the founder of the Bhaṭṭa school of Mīmāṃsā.)

Maṇḍana Miśra (*c*.700?) *Sphoṭasiddhi (Demonstration of the Sphoṭa)*, ed. and trans. M. Biardeau, *La Démonstration du Sphoṭa*, Pondicherry: Institut Français d'Indologie, 1958.(A response to Kumārila's *Ślokavārttika*.)

Matilal, B.K. (1988) '*Śābdabodha* and the Problem of Knowledge-Representation in Sanskrit', *Journal of Indian Philosophy* 16: 107-22.(On *śābdabodha*, mentioned in §2 of the present entry.)

Matilal, B.K. (1990) *The Word and the World: India's Contribution to the Study of Language*, Delhi: Oxford University Press.(Discusses various issues in Indian linguistic philosophy.)

Padoux, A. (1990) *Vāc: The Concept of the Word in Selected Hindu Tantras*, Albany, NY: State University of New York Press.(See especially chapter 5, on the meaning attributed to individual sounds.)

Pāṇini (4th century BC) *Aṣṭādhyāyī(Eight Chapters)*, ed. and trans. O. Böhtlink, Leipzig, 1887; repr. Hildesheim/New York: George Olms, 1997.(The earliest surviving Sanskrit grammar.)

Patañjali (2nd century BC) *(Vyākaraṇa-)Mahābhāṣya (Great Commentary)*, ed. F. Kielhorn, 1880-5, 3 vols; 3rd edn, ed. K.V. Abhyankar, Poona: Bhandarkar Oriental Research Institute, 1972.(A commentary on Pāṇini's grammar and Kātyāyana's *vārttikas*.)

Raja, K.K. (1969) *Indian Theories of Meaning*, Madras: The Adyar Library and Research Centre; 2nd edn, 1977. (Useful but generally nonhistorical survey of fundamental concepts.)

Scharfe, H. (1977) *A History of Indian Literature*, vol. 5, fasc. 2, *Grammatical Literature*, Wiesbaden: Harrassowitz.(Historical survey of technical linguistic literature; pays little attention to philosophical issues.)

Language, innateness of

Is there any innate knowledge? What is it to speak and understand a language? These are old questions, but it was the twentieth-century linguist, Noam Chomsky, who forged a connection between them, arguing that mastery of a language is, in part, a matter of knowing its grammar, and that much of our knowledge of grammar is inborn.

Rejecting the empiricism that had dominated Anglo-American philosophy, psychology and linguistics for the first half of this century, Chomsky argued that the task of learning a language is so difficult, and the linguistic evidence available to the learner so meagre, that language acquisition would be impossible unless some of the knowledge eventually attained were innate. He proposed that learners bring to their task knowledge of a 'Universal Grammar', describing structural features common to all natural languages, and that it is this knowledge that enables us to master our native tongues.

Chomsky's position is nativist *because it proposes that the inborn knowledge facilitating learning is* domain-specific. *On an empiricist view, our innate ability to learn from experience (for example, to form associations among ideas) applies equally in any task domain. On the nativist view, by contrast, we are equipped with special-purpose learning strategies, each suited to its own peculiar subject-matter.*

Chomsky's nativism spurred a flurry of interest as theorists leaped to explore its conceptual and empirical implications. As a consequence of his work, language acquisition is today a major focus of cognitive science research.

1 The development of Chomsky's nativism

In a review of Skinner's *Verbal Behavior*, Chomsky ((1959) rejected the behaviourist view that mastery of a language, or 'linguistic competence', consists in complexes of 'dispositions to verbal behaviour' instilled in our minds by 'operant conditioning' during childhood (see Skinner, B.F.; Behaviourism, methodological and scientific). First, children neither need nor receive the careful linguistic training that the behaviourist acquisition-theory requires. Second, our use of language is both *stimulus-independent* (not determined by the inputs we receive) and *productive* (we can utter and understand indefinitely many novel sentences). Hence competence must be more than a congeries of 'verbal habits'. Both language acquisition and linguistic competence involve complex systems of psychological states and processes.

But how to characterize these states and processes? In the early 1960s, Chomsky's 'generative' approach to linguistics challenged the 'taxonomic' approach of American structuralists such as Bloomfield and Harris (see Structuralism in linguistics §4). He urged that linguists abandon the structuralist aim of elucidating 'discovery procedures' (that is, mechanical methods which, when applied to a corpus of sentences in a language, will deliver a grammar descriptive of that language) and focus instead on the development of rigorous, formalizable syntactic theories for particular languages. The linguist's goal should be to develop *generative* grammars for natural languages, that is, sets of rules that will produce all and only the sentences in a language, together with their syntactic-structural descriptions (see Syntax §3).

Much of Chomsky's work in linguistics has aimed at refining the generative programme. Of particular interest here, however, is his contention that a grammar's usefulness outruns its ability to characterize a language. Its real importance, he argued, is psychological. Since speakers' grammatical intuitions (their judgments as to the well-formedness of sentences) constitute the data for a grammatical theory, that theory is in some sense descriptive of what speakers know about their language. Thus, Chomsky inferred, grammars supply (partial) accounts of the psychological structures underpinning language use, so providing a timely replacement for the ousted behaviourist account of linguistic competence. (Chomsky distinguishes theories of linguistic competence from theories of linguistic performance. Competence theories describe our capacity for language in a way that abstracts from the question of how that capacity results in the production and comprehension of sentences. Performance theories explain how the system of knowledge postulated in a competence theory interacts with other psychological systems such as memory, perception and the articulatory systems, in order to produce actual linguistic behaviour.)

But if grammars are competence theories, the question of how language is *learned* becomes acute. Normal children master language by about the age of eight. On Chomsky's view, this mastery entails, among other things, the

'internalization', 'cognizance' or 'tacit knowledge' of their language's grammar. But a grammar is a highly complex body of syntactic, semantic, morphological and phonological rules, employing concepts and constructs far removed from experience. How do children internalize so complex a grammar at so tender an age, with apparently no conscious effort or instruction?

Chomsky argues that children's acquisition of grammar is explicable only on the assumption that a substantial portion of the knowledge eventually attained is innate. On his view, a 'Universal Grammar' (UG), which specifies information about 'linguistic universals' (features common to all natural languages), is embodied in the language-learning mechanism itself. Thus, children need not *learn*, for instance, that declarative sentences must have a noun phrase as their subject: that fact they bring with them to the learning task. All they need do during learning is determine those facts (such as that, in English, the subject precedes the verb) that are *specific* to their language.

Chomsky is notable in that he does not rest content with the claim that we have innate knowledge of UG. Unlike his nativist predecessors, such as Plato, Descartes and Leibniz (see Nativism §2), he has offered two concrete proposals as to how that knowledge is realized.

Up to the late 1970s, he defended a 'hypothesis-testing' model of learning according to which children unconsciously project hypotheses about the grammar of their language, testing them against the data provided by experience. Their innate knowledge of UG, on this view, takes the form of constraints on the kinds of hypotheses they can entertain: they must be formulated in a particular language; they must have a certain form; and, in cases where the evidence is insufficient to arbitrate between two competing hypotheses, the language-learning mechanism itself will supply a decision.

More recently, Chomsky has embraced a picture of language learning as involving 'parameter-setting'. On this view, which is as yet only partially developed, the principles of UG are encoded in the mind of the neonate, each principle containing one or more variables or 'parameters'. A parameter's possible values are extremely limited and are fixed by experience, different parameter-settings resulting in competencies in different languages. The 'null subject parameter', for example, governs whether one may omit the grammatical subject of a spoken sentence. Exposure to a language, like Spanish, where subjects may be omitted, sets the parameter one way; exposure to a language, like English, where subjects must be phonologically realised, will cause it to be set differently. Thus, children's linguistic experience does not function as *evidence* for or against their grammatical hypotheses. Rather, and in much the same way as their hormones trigger the bodily changes that take place at puberty, exposure to the language of their community 'triggers' the mental changes that eventually result in linguistic competence.

Evaluating the merits of these models is in large part an empirical task and is outside the scope of this article. We will consider instead the case for linguistic nativism itself; that is, for the view that language-learning requires a task- or domain-specific mechanism of some kind.

2 The poverty of the stimulus

The commonest motivation for nativism is the argument from the poverty of the stimulus (POS). It contends that there is a 'gap' between the rich stock of linguistic knowledge that competent speakers acquire and the meagre supply of information about language that their experience during learning provides. This gap is so wide, the nativist contends, that no child endowed only with an empiricist-style, general-purpose learning mechanism could hope to bridge it. Language learning, therefore, requires a special-purpose device, one that embodies at the start some of the information about language that is ultimately attained.

This basic argument is elaborated in two ways. There are several a posteriori versions, due mainly to Chomsky; and a rather more a priori version with roots, seemingly, in Formal Learning Theory, a branch of mathematical linguistics.

A posteriori POS arguments are based on specific empirical claims about the inputs to and outputs of learning. In *Reflections on Language* 1975: 31 and following), for example, Chomsky claims that (1) competent speakers know the auxiliary fronting rule for question-formation; and that (2) the information available to children does not reliably distinguish the correct rule from incorrect but equally plausible alternatives. Hence, he infers, the

learner's choice of the correct rule is driven not by any general principles of theory-construction (such as the empiricist envisages), but is due rather to the fact that the acquisition device itself is biased towards grammatical rules of certain kinds, namely those conforming to UG.

Note that a posteriori POS arguments rely on specific empirical assumptions about the nature of linguistic competence and the data accessible to children. If it were discovered, for example, that speakers had *not* in any interesting sense internalized the rule in question, or that data disconfirmatory of incorrect rules were readily available, Chomsky would have failed to locate a gap - and hence to have supported his nativism. Hence, the success of a posteriori arguments turns on the degree to which their factual premises can be empirically sustained. But while some steps have recently been taken towards furnishing the supporting (for example, developmental) data required by arguments like this, their overall bearing on the nativism-empiricism controversy remains undecided.

In contrast to the preceding, the a priori POS argument (often called the Logical Problem of Language Acquisition) seeks to infer the existence of a gap from a consideration of the logical structure of the acquisition-task itself. Take some state in the learning process at which a child has acquired a grammatical hypothesis. Call the set of sentences (the language) generated by their hypothesised grammar H. Call that generated by the correct grammar (the target language) L. Then the learner is in one of five situations: (1) H is disjoint from L; (2) H overlaps L; (3) H is a subset of L; (4) H is a superset of L; (5) H is equivalent to L.

Situation (5) represents the end point of the learning process: the child has learned the target language. Situations (1), (2) and (3) are relatively unproblematic: the learner will be forced to revise their hypothesis if they hear a sentence in L that is not in H. Thus, they can converge on the correct grammar by exposure to 'positive evidence', evidence (as is provided by someone's uttering a sentence) that a given string of words is a sentence of the target language. But, crucially, if they are in situation (4) - if their grammar 'overgenerates', that is, generates all of L and also some sentences that are *not* in L - they cannot discover *this* by exposure to further sentences of the target language. For, every sentence of L is equally a sentence of H. What this learner needs is the information that a sentence that is generated by their grammar - a sentence in H - is *not* a sentence of L; only then can they 'shrink' their hypothesis so that it converges on the correct one.

This latter kind of evidence, that a sentence is *not* in the target language, is called 'negative evidence'. And case (4) is of interest because negative evidence is in general unavailable to children. *Explicit* negative evidence is scarce: children are not provided with lists of ungrammatical sentences; their errors are rarely corrected; nor do parents typically notice their own ungrammatical utterances. And *indirect* negative evidence is hard to find too. The non-occurrence of a sentence in the data, for example, is not negative evidence. For, there are infinitely many sentences of L that *de facto* will never be uttered. Hence from the mere fact that they have not heard a certain string of words in the data to date, the child cannot infer that it is not a sentence of the target language.

Thus learners who find themselves in situation (4) have no systematic access to evidence that would enable them to correct their overgeneral hypotheses. Which means, since children *do* eventually converge on the correct grammar, that they must never be in situation (4); the learning mechanism must be constrained such that they *never* hypothesize a grammar that is 'too large'.

Nativists contend that the general-purpose constraints on hypothesis-formulation envisaged by empiricists are insufficient to ensure that overgeneration will not occur. Hence, they conclude, the space of possible grammars must be constrained by further, linguistically-specific principles such as are provided by UG: the poverty of negative evidence in the data requires a nativist approach to language acquisition.

The a priori POS argument is at first sight much more compelling than its empirically undersupported a posteriori relatives. It is not, however, without points of weakness. First, one may question whether negative evidence is really as scarce as it seems. For example, while the non-occurrence of a particular sentence in the data may not be evidence that it is not grammatical, perhaps the repeated non-occurrence of certain syntactic forms could be evidence that sentences of that form are not grammatical. (That there must be at least *some* negative evidence available is evident from the fact that children do, as a matter of fact, recover from overgeneral grammatical hypotheses: they learn, for example, that you do not *always* add /d/ to the stem to form the past tense of a verb).

Second, one may question the nativist's contention that general-purpose constraints are insufficient to prevent a

learner's making incorrigible overgeneralizations. Not only is this contention to my knowledge unargued, there is reason to believe that in at least *some* cases, general-purpose learning strategies *are* efficacious in the face of what seem to be identical 'negative evidence' problems. For, negative evidence - evidence as to what things are *not* - is *quite generally* unavailable. When we learn what cars are, for example, we are not systematically informed about the countlessly many things that are not members of that class. Yet, and despite this paucity of negative evidence, we do manage to converge on the correct hypothesis about 'car-bonnet'. So, assuming that it is implausible to postulate a task-specific 'automotive faculty' that facilitates our learning in this domain, it must be the case that whatever general-purpose learning strategies we possess are in at least some contexts able to function successfully in the (near) absence of negative evidence.

In sum, arguments from the POS alone cannot compel nativism about language. We need additional reasons to think that there is something *special* about language acquisition - features it does not share with learning about, say, cars - such that nativism is plausible in the former case, while being clearly implausible in the latter.

3 Other arguments for nativism

Language acquisition has been held, by Chomsky and others, to be special in a number of ways that support linguistic nativism. Examples of these claims, and of the empiricist responses to them, are:

(1) *Species specificity.* Only human beings, so far as we know, speak a language. We thus possess a special biological affinity for language, an affinity which may amount to an innate language-learning faculty. Response: only human beings build cars, too, but that does not make for an innate automotive faculty. What our species possesses that others do not is general intelligence - an ability to learn from experience in a wide variety of contexts.

(2) *Critical period.* Language acquisition displays 'critical period effects': the ability to learn language declines drastically after puberty. This resemblance to other biologically-controlled developmental processes suggests that the acquisition of language too is determined more by our genes than our environment. Response: perhaps *all* learning declines after puberty. There are no data supporting the view that language-learning is special in this regard.

(3) *Facility of acquisition.* Language is acquired with remarkable speed and ease at a very early age. Processes distinct from those underpinning other kinds of learning are therefore at work. Response: given that we lack anything to compare language learning with, it is unclear that it is either notably fast, or especially easy.

(4) *Linguistic universals.* There are deep syntactic and semantic similarities among all known natural languages. Since there is no evidence to support the view that now-current languages descended from a common ancestor, these commonalities must be consequences of the way the language faculty is structured. Response: of course languages are shaped by the minds that use them. But linguistic universals may be due to constraints on what can be learned via a general-purpose learning strategy, rather than arising out of the structure of an innate language faculty.

(5) *Linguistic complexity.* Linguists, using *exactly* the 'general principles of theory-construction' beloved of the empiricist, spend entire careers laboriously constructing grammars for tiny corners of a language. But children can learn *all* of the grammar of *any* language simply by hearing it spoken around them. Surely some special faculty is at work here? Response: this argument assumes that there *is* an interesting sense in which speakers know - hence must learn - the grammar of their language. This assumption, however, may be challenged. First, there are conceptual difficulties involved in understanding what it is for a speaker to know a grammar. Thus, it is unclear whether children's ability to learn one is remarkable or not. Second, it has proved difficult to find empirical confirmation for the claim that grammars are implicated in any straightforward way in on-line linguistic processing. Thus it is unclear in what sense, if any, grammars need to be learned. And finally, one might question whether the linguist's methodology is appropriate for investigating the psychology of language use and understanding. Thus, while it is remarkable that the linguist's task in characterizing the language the child knows is dauntingly complex, this is of unclear relevance to the child's task in learning it.

In conclusion, Chomsky's claim that knowledge of language is largely innate has renewed the interest of linguists, philosophers and psychologists in the problem of explaining language acquisition. It remains, however, an open empirical question whether that explanation requires the postulation of a special-purpose learning mechanism embodying innate knowledge of language; and, if it does, what the nature of that mechanism and that knowledge

is.

See also: Chomsky, N.; Innate knowledge; Language, philosophy of; Semantics §1

FIONA COWIE

References and further reading

Chomsky, N. (1959) 'Review of Skinner's *Verbal Behavior*', *Language* 35: 26-58.(Chomsky's critique of behaviourism; discussed in §1.)

Chomsky, N. (1975) *Reflections on Language*, Glasgow: Fontana/Collins.(Seminal presentation of *de facto* POS argument (in §2) and the hypothesis-testing model of learning discussed in §1.)

Chomsky, N. (1988) *Language and Problems of Knowledge: The Managua Lectures*, Cambridge, MA: MIT Press. (Chomsky's current views about the innateness of language. Discussion of parameter-setting (in §1). References to earlier works.)

Hornstein, N. and Lightfoot, D. (eds) (1981) *Explanation in Linguistics: The Logical Problem of Language Acquisition*, London: Longman.(Sympathetic exploration of a priori POS argument; discussed in §2.)

Katz, J.J. (ed.) (1985) *The Philosophy of Linguistics*, Oxford: Oxford University Press.(Chomsky's nativism is a major focus of this collection.)

Matthews, R.J. and Demopoulos, W. (eds) (1989) *Learnability and Linguistic Theory*, Amsterdam: Kluwer. (Implications of the a priori POS argument (discussed in §2) for the theories of language learning.)

Newmeyer, O. (ed.) (1988) *Linguistics: the Cambridge Survey*, vol. 3, *Language: Psychological and Biological Aspects*, Cambridge: Cambridge University Press.(Empirical approaches to language acquisition. More up-to-date than Wanner and Gleitman (1982), but less focused on the nativism issue.)

Piattelli-Palmarini, M. (1989) 'Evolution, Selection and Cognition: From "Learning" to Parameter Setting in Biology and in the Study of Language', *Cognition* 31: 1-44.(Introduction to parameter-setting discussed in §1, with references.)

Pinker, S. (1979) 'Formal Models of Language Learning', *Cognition* 7: 217-82.(Survey of formal learning theory literature discussed in §2.)

Pinker, S. (1994) *The Language Instinct*, New York: William Morrow.(An excellent and highly readable summary of the case for linguistic nativism.)

Sells, P. (1985) *Lectures on Contemporary Syntactic Theories: An Introduction to Government-Binding Theory, Generalized Phrase Structure Grammar and Lexical-Functional Grammar*, Stanford, CA: Center for the Study of Language and Information.(Introduction to formal syntax, including a chapter on generative grammar relevant to §1.)

Wanner, E. and Gleitman, L.R. (eds) (1982) *Language Acquisition: The State of the Art*, Cambridge: Cambridge University Press.(Collection of psychologists' writings on language acquisition, with special emphasis on nativism.)

Language, medieval theories of

*A great deal of theorizing about language took place in western Europe between 1100 and 1400. The usual social context of this theorizing was the teaching of grammar, logic or theology. Rhetoric was traditionally counted as one of the language disciplines (*scientiae sermocinales*) together with grammar and logic, but in practice it received little attention. Medieval thinkers produced a vast literature on aspects of linguistic theory, but they did not write books with such titles as 'A Theory of Language'. The theories that have come down to us today have been reconstructed from a large number of sources, even when they are attributed to a single person.*

Although the medieval writers on language were very innovative, they owed some key ideas to ancient Greek and Latin authors, for example: (1) words acquire their meaning by an act of 'imposition' when a sound is chosen as the label of some thing; (2) there are three key ingredients in signification: the word, a concept and the thing signified; (3) concepts can be thought of as mental words; (4) the grammaticality of a sentence cannot be explained purely in terms of morphology; and (5) words have different contents when used as predicates of creatures and when predicated of the Creator.

The medieval thinkers disagreed as to whether words signify things directly or only through concepts. The latter view ran the danger of making concepts a screen between language and reality, but it had the advantage of being able to explain why different words can signify the same object without being quite synonymous. Many thought that general terms signify universal things, but there were also nominalist schools which held that the general terms themselves are the only universal things. Fourteenth-century nominalists located universality in concepts, also called mental terms, and only secondarily in spoken words. The language of thought had gained priority over that of speech.

The theory of grammar known today as modism was developed in the thirteenth century. This theory assumed that there are only two contributors to the sense of an expression, lexical meaning and grammatical features, and that only the latter belong to the province of grammar, which thus became a purely formal science that could claim applicability to all languages irrespective of their surface differences. The problem with modism was that it had no tools for dealing with even slightly deviant, yet intelligible expressions. Thus it would have to reject a statement such as 'the crowd are rushing' because of the lack of concord of number.

A number of medieval scholars focused on the various forms of metaphorical language and ambiguity, which were more appropriate for dealing with deviant expressions of the type mentioned above. It was realized that speaker, listener and context must be taken into account in order to explain how words can communicate something different from their primary sense, and how it is possible for a listener to grasp the intended sense of an ambiguous message. One motivation for this study was a need to understand how theological discourse functions. Theology also lay behind a heated debate about the ontological status of the meanings of propositions, and sacramental theology joined grammar in developing a notion of performative locutions.

The study of syntax yielded many new concepts, including those of government and dependence. Much less work was done on the evolution of languages, but Roger Bacon and Dante did offer some perceptive observations. Though never creating fully fledged artificial languages, logicians did develop a semi-artificial Latin.

1 Sources

The most important source for grammar in the Middle Ages was Priscian's *Institutiones grammaticae (Institutions of Grammar)*, a comprehensive Latin morphology and syntax composed in the early sixth century AD. Priscian not only describes and prescribes but also reasons about linguistic problems, using a theoretical framework borrowed from the second-century Greek grammarian Apollonius Dyscolus. Priscian transmitted to posterity the notion that the grammaticality (*congruitas*) of a sentence depends on a proper fit among the conceptual elements signified by its constituent words. Also important were his classifications of words, including the eight-fold partition of constituents of sentences as *partes orationis* (parts of speech). These eight parts were known from other sources, but Priscian provided the best discussion of the subject. Matters left untouched by Priscian, such as tropes and figures, were treated in Chapter 3 of Donatus' *Ars Maior*, composed in the fourth century. Under the name of *Barbarismus*, this chapter became a focus around which investigations of deviant expressions developed.

In logic, Aristotle was the main source of inspiration. His *Sophistical Refutations* made it clear that an argument does not always carry its validity on its surface, and his explanations of why a bad argument may look good gave impetus to the development of studies in semantics, and to a lesser degree in the pragmatics of linguistic communication. In *Peri hermeneias (On Interpretation)* 1, Aristotle briefly sketches which relations hold between corresponding items in writing, speech, thought and external reality. According to the standard interpretation of this text, Aristotle says that written words signify spoken ones, which in turn signify concepts, while concepts signify things. This reading can support the notion that truth, validity and their opposites are not primarily properties of spoken or written sentences and arguments, but of their mental counterparts. If the Aristotelian level of concepts is equated with that of Priscian, grammatical and logical correctness (congruity and truth) have the same primary bearers, that is, mental propositions whose 'words' are concepts. Aristotle's definitions of noun and verb in *Peri hermeneias* 2-3 differ from those of the Latin grammatical tradition, and this tension between grammar and logic created fruitful discussion. In *Categories* 1, Aristotle introduces the notion of paronymy or denomination: for example, the just (person) is denominated from justice. From this seed, theories of primary and secondary signification developed (to continue with the example, 'just' primarily signifies justice and secondarily signifies a person).

The logical works of Boethius provided the notion of imposition (*impositio*). Following Porphyry, Boethius taught that words acquire their meaning by an act by which a human 'impositor' decides to use a certain sound to name some thing. He distinguished a first imposition, by which features of the physical world acquire names ('tree', 'green', 'three cubits long' and so on), from a second one by which names acquire names ('noun', 'verb' and so on). He further suggests that a word such as 'species' is a name of names, so that 'man is a species' is a statement not about man but about the word 'man'. This suggestion helped to fuel early nominalism.

In theology, some essays by Boethius were influential until about 1200, but the principal authority was always Augustine. These two authorities inculcated in medieval writers the conviction that ordinary words do not have their ordinary meanings when used of God. For God to be high and just is not the same as his having big quantity and the quality of justice. God in his absolute simplicity cannot be analysed in terms of the Aristotelian categories. This doctrine was particularly important in the twelfth century when, according to a widely accepted analysis, ordinary predication was thought to consist in the ascription of a certain form to some bearer or bearers. Any such form was supposed to belong to some Aristotelian category. Thus, 'Socrates is a man' ascribes the substantial form 'humanity' to Socrates, 'Socrates is just' ascribes the quality of justice to him, and so on. Since the truth of 'God is just' was uncontroversial, the Augustinian-Boethian doctrine forced medieval writers to deny one or both of the claims that (a) a predicate term always signifies some categorial form, and (b) in propositions of the type '*X* is *Y*', the form signified by *Y* is always attributed to the referent(s) of *X*. Usually, some combination of the two solutions was advanced. In the first place it was claimed that 'divine terms' ('God' in particular) never signify categorial forms; in the second place it was claimed that true propositions of the type 'God is *Y*' never ascribe categorial forms to God, but either indicate a sort of total identity between God and *Y* or actually ascribe a categorial form to his creation, so that 'God is just' may be paraphrased as 'there is justice in the created world, and it derives from God'.

Augustine (*De trinitate* XV) also contributed to the theory of mental speech his notion of a soundless 'speech in the mind' (*verbum mentis*) prior to any language in the phonic medium. He further gave a definition of a sign as that which, besides presenting itself to the senses, also makes the mind think of something other than the sign itself. In somewhat different wordings, this definition also occurs in Augustine's *De doctrina christiana* 2.1.1 and *De dialectica* 5. The last-named work had few readers; one, Roger Bacon, developed from Augustine's work an unusual awareness that signification is a three-term relation involving not only sign and *significatum* but also someone to whom the sign signifies the *significatum*.

The whole patristic tradition of Biblical exegesis provided another rich source for medieval speculation about linguistic matters, not least about metaphorical language.

2 Words, concepts and things: introduction

Words were said in the Middle Ages to signify (*significare* or *designare*). A special terminology was developed for reference, but it remained a matter for debate what sort of entities significates actually were.

Few writers of the period were very concerned about written words, which they considered as mere 'stand-ins' for spoken words, although Pierre d'Ailly pointed out that a written message can be understood directly without recourse to sound. Concerning spoken words, late ancient sources suggested that they signify things through concepts. Typological interpretation of Biblical passages introduced the further notion of words signifying things which in turn signify other things, but this model was relevant only in Biblical studies. Some, such as John Buridan, took the word-concept-thing model to mean that words signify concepts which in turn signify things, so that words signify things only thanks to the transitivity of signification. Others, notably several late-thirteenth-century authors as well as William of Ockham in the fourteenth century, took it in the sense that for words to have the relation of signification to things, they must also be related in a certain way (subordinated) to concepts.

Those who took words to signify things directly usually did so because they felt that otherwise the concepts would become an impenetrable screen between language and external reality. If concepts were what our words signify, then it seemed we would always be talking about mere psychological entities. Buridan preferred the alternative model, *inter alia* because it can explain how two words can have the same extramental referents without being strictly synonymous. It suffices that the two words have different concepts for their direct significates.

The link between word and significate was generally held to be conventional (*ad placitum*), as Aristotle had said, but this was understood more in the sense that the link was the result of a free choice than in the sense that it depended on agreement between the users of the language. This link was supposed to have been established by 'the impositor', a figure who can trace his ancestry through Boethius and Porphyry back to the 'giver of names' in Plato's *Cratylus*. Some authors seem to have understood the 'impositor' very literally as some wise man who at some time had decided what things were to be called. Others took 'the impositor' to be a collective designation of the people who had introduced new words in course of time, so that in practice 'word *X* was imposed to mean *Y*' equalled 'the traditional meaning of *X* is *Y*'. Still others distinguished between the original wise impositor(s) who devised the phonology and other grammatical features of the language and laid down a basic vocabulary, and later users of the language who added to the vocabulary through new acts of imposition while not introducing major systematic changes.

Boethius operated with a first imposition to produce words for natural phenomena, and a second one to introduce words predicable of words (such as 'noun' or 'trisyllabic'). Early nominalists included terms such as 'species' among those of the second imposition, but after the thirteenth century the vocabulary of logic was more often described as words of second intention, 'second imposition' being reserved for the vocabulary of grammar. Neither the distinction according to imposition nor that according to intention was designed to cover the whole of a language's vocabulary, but only such items as are predicable (*categore(u)mata*). The understanding of what sort of entities first and second 'intentions' are changed over time. To Ockham, a word of the second intention is one that signifies mental *intentiones* (concepts), which are natural signs of other things (the individuals falling under the concept). By contrast, words of the second imposition signify things which are conventional signs, namely words.

While most medieval writers ascribed freedom of choice to 'the impositor', they also assumed that sounds had not been selected at random. Thus there was a justification for etymology, the art of revealing which properties of some thing the impositor paid special attention to when selecting a sound to name it. For example, many found it probable that the Latin word for 'stone', the accusative form of which is *lapidem*, had been chosen because it is characteristic of stones to hurt one's foot, and in Latin 'hurting the foot' is *laedens pedem*, which is somewhat similar to *lapidem*. Moreover, this aggressive feature of stones could be alleged as a reason for the masculine gender of *lapidem*. Theologians commenting on Genesis 2: 19-20, where Adam appears as the first giver of names, stressed that he selected appropriate names on the basis of his understanding of the nature of the things to be named. This conventionality of language was particularly emphasized by a number of fourteenth-century thinkers, who felt that the previous age had tended to produce overly rich ontologies by assuming some real distinction corresponding to every linguistic distinction they saw.

The act of imposition could be understood as one of uniting the material aspect of a word (sound) with a formal aspect (informational content). Thirteenth-century scholastics worried about how this could be possible. One proposal was that just as the formal features of a material thing can be transferred to the mind as an immaterial likeness (species), so the intelligible species can be communicated to other people by means of a material vehicle,

the mind impressing a sort of copy of the relevant species on any particular token of a type of the sound conventionally associated with some sort of thing. Roger Bacon took this mechanism to be the explanation of magic and, apparently, of ordinary persuasion. He thought that if the speaker's mind and body cooperate to emit sound carrying some species, the species may act as a force (*virtus*) capable in favourable conditions of producing changes in both material and spiritual things. However, many thinkers of the next century were to deny material words the ability to carry information; instead, they located the information in the concepts to which words correspond.

3 Words, concepts and things: nominalism

Shortly before 1100, some philosophers began to define universals as words (*voces*) or names (*nomina*), and to consider predication a linguistic phenomenon: words are predicated of words, not things of things. Logic thus became a genuinely linguistic discipline (see Logic, medieval). Porphyry's five predicables (genus, species and so on), Aristotle's ten categories (substance, quantity and so on), and Boethius' topical differences (genus, opposite and so on) could all be viewed as names applicable to certain classes of words, that is, as words of the second imposition.

Early representatives of this 'vocalism' apparently did not make it sufficiently clear that the voces that are the bearers of universality are not mere articulated sounds without regard to their meaning, and thus they laid themselves open to the objection that, in their theories, the Greek *anthropos* and the Latin *homo* would be two different universals even though they both mean 'man'. Likewise, *homo* uttered at one time would not be identical with the equiform sound uttered at another time. The supposed universal would thus dissolve into a host of individual sounds.

This problem in early 'vocalism' was solved when Peter Abelard, the father of a school of *nominales*, made it clear that the criterion of identity for a universal was its signification: several voces with a shared signification make up one *sermo*, he held, and *sermones* are the bearers of universality. The signification of a universal term consists first in naming (picking out) individual things, and second in a relation to mental acts (*intellectus*). The individuals that a universal (such as 'man') picks out share a circumstance (*status*), for example *esse hominem* (being a man), but a *status*, Abelard held, is no thing. Some of his followers seem to have taken a *status* to be what a universal signifies; thus terms acquired significates of a type similar to those of whole propositions, for 'Socrates is a man' was supposed to signify the *dictum* that-Socrates-is-a-man (*Socratem esse hominem*) (see Nominalism).

About 1200, nominalism almost disappeared and only returned after 1310 with William of Ockham and John Buridan as its leading figures. Both the latter operated with three types of language, written, spoken and mental, and of these the mental language was considered logically most important. According to Ockham, terms of the written language are 'subordinated' to terms of the spoken language, and they in turn are subordinated to terms of the mental language. A mental term is a natural sign of some set of individuals. Its subordinate spoken and written terms are conventional signs of exactly the same individuals, so signification always links a term to individual things. Buridan sometimes uses Ockham's terminology, but more often he assumes that written terms signify spoken terms which in turn signify mental terms, which finally signify individual things, the 'ultimate significates'.

To both Ockham and Buridan, mental terms are concepts, and they are the primary bearers of universality. Oral and written terms are so derivatively. Similarly, 'true' and 'false' are primarily predicates of mental propositions. In Ockham's mature theory, a concept is identical with an act of intellection, and its ontological status is therefore that of an accidental characteristic of a soul. The ability of a concept C to signify every member of some set S derives from a combination of (1) a causal history reaching back from C to some member x of S encountered in direct cognition (intuition), (2) an isomorphism between C and x, and (3) a maximal similarity between x and any other member of S.

Ockham's mental language was almost a duplication of spoken language, but sometimes, at least, he thought of it as a language purified of the ambiguities and redundancies of its spoken counterpart and held that spoken language has two strictly synonymous terms, mental language has only one. Similarly, he claimed that grammatical features of spoken language lack mental counterparts if they are irrelevant to the truth of a proposition. Thus mental language does not have grammatical gender, but it does have tense.

Ockham and Buridan assumed multiple denotation for general terms whether they be in the subject or the predicate position, and co-reference of subject and predicate term was their fundamental criterion of truth for a categorical proposition. According to them, an affirmative categorical proposition is true if and only if the subject and the predicate term each refer to (*supponit pro*) at least one individual, and any individual to which the subject term refers is also referred to by the predicate term. The occurrence of an empty term like 'chimera' renders any affirmative proposition false and any negative proposition true (see Propositions, sentences and statements).

Thirteenth-century thinkers typically took signification to be a relation linking words with universal concepts, common natures or the like. The technical term for the relation between word and individual things was 'supposition' (see §12). For fourteenth-century nominalists, however, both signification and supposition terminate in individuals, but a word supposits only when occurring as a term in a proposition whereas it signifies even when uttered in isolation. Generally, a word's *supposita* will be a subset of its significates. The exception is the case when a word supposits for itself (or equiform words), whether in the mental, the spoken or the written variant, such as 'man' in 'man is an absolute concept', 'man is a monosyllabic word' or 'there are three letters in man'. In such cases, the term was said to be used non-significatively.

Whereas the verb's tense and modality define the relevant set of *supposita*, quantifiers affect the way a term supposits and thus the way to spell out the truth conditions of the proposition by means of the technique known as exposition. Exposition replaces the original proposition by an equivalent conjunction or disjunction of propositions, or by one proposition with a conjunct or disjunct subject or predicate. Thus 'every man runs' → 'this man runs, and that man runs, and so on', 'some man runs' → 'this man runs, or that man runs, and so on'.

In the early fourteenth century, the dominant grammatical theory was based on the notion that words possess 'modes of signifying' from which they derive their grammatical properties. It took some time for nominalist philosophers to work out an alternative, but by 1400 a new grammar had developed which took mental language to be the primary bearer of grammatical as well as of logical predicates. Mental terms were said to signify 'nominally' or 'verbally' the grammatical categories built into the concepts. Well-formedness was claimed to belong to spoken strings of words only thanks to their 'subordination' to grammatical strings of mental terms, not thanks to any grammatical properties inherent in the spoken words themselves, for how could sounds bear such properties, which obviously have an origin in the mind? Proponents of such views on grammar included Albert of Saxony, Pierre d'Ailly and one Marsilius (possibly Marsilius of Inghen). Some fifteenth-century thinkers also accepted modes of signifying, but held that they resided in concepts, not in words.

One strange result of late medieval nominalism was that thought was often considered to be more genuinely language than speech.

4 Words, concepts and things: realism

Taking universality to be a predicate only of concepts and/or words was not the prevailing view in the Middle Ages; more often, universality was held to reside somehow in things, though requiring a mental operation to be brought out into the open. This is not the place for a history of realist theories (see Realism and antirealism), but two points need to be mentioned. First, the naïve Platonism in which universal Forms, residing in a world of their own, are the principles of being and of understanding for perishable things, while being also the significates of general terms, never had currency during this period. Second, probably the most influential variant of realism was one that arose in the thirteenth century and owed much of its inspiration to the Persian philosopher Avicenna (see Ibn Sina). While accepting the reality of universals, this theory did not make them the significates of general terms. Instead, it held that such words signify 'common natures' transcending the distinction between universal and particular, existence and non-existence. As the common nature was generally identified with quiddity and essence, it was considered a principle of the understanding and being of things, but not one that would bestow either universality, particularity or existence on them. Particularity and universality were considered different manifestations, as it were, of the same underlying common nature or *res* (thing). Such 'things', the theory held, may have being in many alternative 'ways' or *modi essendi* (modes of being), in a static and in a dynamic way, in a universal or a particular way and so on, and this is reflected in human understanding and language. Thus the distinction between nouns and verbs brings out the distinction between the static and the dynamic modes of being of things.

5 Words, concepts and things: modism

About 1260-80 a grammatical theory of high generality was developed, known as *grammatica speculativa* (theoretical grammar); today the standard term for this theory is 'modism'. The first generation of 'theoretical grammarians' included Boethius of Dacia, Martin of Dacia and John of Dacia. Radulphus Brito and Thomas of Erfurt, both active around 1300, belong to the second generation. Few theoretical advances seem to have been achieved after about 1310, though the works of Martin and Thomas continued to be used for about two centuries.

The term *modus significandi* (way/mode of signifying) had been in use since the twelfth century, typically to explain why an abstract and a concrete word for the same thing are not interchangeable *salva veritate*. Stephen Langton, for example, claimed that *deus* (God) and *deitas* (godhood) have the same signification and the same referent, and yet the truth of *deus est in lapide* (God is in a stone) does not entail the truth of *deitas est in lapide* (godhood is in a stone) because *deus* and *deitas* do not signify in the same way. The first proposition is true because of God's ubiquity, the second is false because it amounts to a claim that the stone is divine. By the mid-thirteenth century, it had become common to call grammatical properties of words 'modes of signifying', and a distinction had been introduced between a word's 'general' (lexical) signification and the 'special signification' due to its grammatical form.

In a further development, the notion, inspired by Avicenna, of common natures and their modes of being provided a suitable metaphysics for a linguistic theory with 'modes of signifying' as the key concept. It was assumed that the structure of human language reflects a correct understanding of the structure of external reality. Each of the traditional grammatical categories reflects one way of thinking, a *modus intelligendi* (way/mode of understanding) of items of reality - *res* or *natura communis* - and each way of understanding corresponds to one of the ways in which such items are, a *modus essendi* (way/mode of being). For a word thus to belong to some grammatical category means that it signifies some thing in a certain way, as having a certain mode of being: for a word to be a noun means that it signifies some thing in the way things are signified under their static aspect (mode), and for a word to be a verb means that it signifies some thing in the way things are signified under their dynamic mode. The noun *dolor* (pain) and the verb *dolere* (to ache) signify the same *res*, but in different ways.

While the modes of signifying are derived from correct understanding of the modes of real things, once they have been found and have found morphological expression, nothing prevents humans from inventing a word such as 'chimera' and furnishing it with a noun's mode of signifying. That mode of signifying indicates that if reality has a match for the word 'chimera', that thing must have static being. No particular facts about the external world can be deduced from grammar, only some general structural features.

Modes of signifying were divided into two types, *essential* and *accidental*. The essential modes establish the eight parts of speech, while the accidental modes establish moods, tenses, numbers, genders, cases and so on. Like the essential modes, accidental modes are derived from features of reality. Thus in reality things can have an active, passive or neutral way of being; this is the origin of the modes of signifying traditionally called masculine, feminine and neuter gender. A word must not be marked for masculine gender if the intended significate can have only feminine being, for that would make it the sign of another sort of thing than the one intended. However, for a word to be comprehensible it is only required that the mode of being indicated by the mode of signifying is compatible with the actual modes of being of the thing in question. Hence for things neutral with respect to action and passion, the choice of gender is arbitrary.

According to several modists, any particular word-form may be analysed as follows. First, it is a vocal sound (*vox*), endowed with a significative function (*ratio significandi*), one essential mode of signifying and some accidental modes of signifying. Having a significative function makes it into a sign, and having articulated sound (*vox*) for the vehicle of its significative function makes it into the sort of sign called a *dictio*. Having an essential mode of signifying makes it a *pars orationis* (part of speech, or in other words, a constituent of a sentence).

A fundamental, and controversial, thesis of modism was that the significative function linking a particular type of sound to some specific significate (*significatum*, *res*) is irrelevant for the grammaticality of constructions. Grammaticality is the result of licit combinations of modes of signifying. Thus {[+ noun + substantive + plural + nominative] + [+ verb + plural + third person]} is a licit combination, and any two words having those modes of signifying make a grammatical construction, no matter what their lexical meanings are. Thus 'stones speak' is as

correct as 'men speak'. (In the schema shown above, the terms 'verb' and so on replace the cumbersome names of the corresponding modes of signifying.)

Modists assumed that external reality is accessible to human understanding, and that whatever can be understood can be expressed. Different peoples may choose different ways of encoding their understanding of reality in sound, but reality is the same for all and is understood in the same way by all peoples. Consequently, all languages must be structurally identical; they must all have the same grammar, surface differences notwithstanding. Greek (the modists wrongly assumed) has lexicalized 'number' as far as nouns are concerned and indicates it by means of special words called articles instead of using endings as in Latin, but what matters is that both languages can express 'number'. If the underlying grammatical system is the same for all languages, it follows that there is total translatability between them, and modists thought as much. Indeed, they found empirical evidence for the thesis in the well-known fact that logic was first developed in a Greek context and written down in Greek, then successfully translated into Latin. If logic is international and translatable, then surely nothing is untranslatable.

Among medieval thinkers, the modists presented the most articulate theory of the universality of grammar. Fourteenth-century opponents of modism, such as John Aurifaber, sometimes explicitly denied the universality thesis.

6 Multiple signification: proper and improper locutions

A pervasive theme in the thought of Anselm of Canterbury is that people often use improper locutions, and that these must be recognized as such and replaced by the proper ones if truth is to be found. A typical example of his use of this technique concerns the argument that for any thing (A) existing at t_1, it was true at $t_{<1}$ that A could come into being, and so any actual thing has a potential predecessor. For Anselm, potency and power were not distinct concepts, and the argument thus became a claim that at $t_{<1}$, A had the power to come into being. It follows that the actual A may have generated itself; thus the created world needs no creator, as it has its own principle of existence. Anselm's way of countering this line of thought is to claim that 'Before A came into being, A could come into being' is a colloquial way of saying 'Before A came into being, God had the power to bring A into being.' The improper locution leaves the false impression that at $t_{<1}$ A is the holder of the power to produce A, when in fact that power is held by God. This distinction between a more and a less proper meaning of words recurs in many guises in medieval thought.

7 Multiple signification: metaphor, allegory and author's intention

Language with one surface meaning and another one hiding beneath the surface was of particular concern to the theologians. For one thing, it was agreed that 'God is just' cannot be interpreted in the same way as 'Peter is just'. For another, it was commonly assumed that not only does the Bible sometimes use expresions that are only meant to be understood metaphorically (for example, 'sprout from the root of Jesse' as a metaphor for Christ), but often a Biblical text carries several senses besides the one apparent on the surface (*sensus historicus*). Thus 'heaven and earth' in Genesis 1: 1 could be said to signify heaven and earth on one level of interpretation, angel and man on another, spirit and flesh on a third. In such cases we are dealing with *allegoria in verbis*, that is, words having another meaning than or besides their *prima facie* meaning.

However, following a suggestion from Augustine in *De trinitate*, medieval writers also assumed an *allegoria in factis* (allegory in deeds), in which the words do not directly have any but their surface meaning, but the thing meant in turn means something else. This occurs in the Bible when words directly signify some event (*factum*) from the time of the Old Covenant, but indirectly also signify some fact belonging to the time of the New Covenant. Thus if the text says that the people of Israel were delivered from Egypt, the words primarily refer to the liberation of the Israelites from the Egyptian captivity; but this event in turn was a sign of the future liberation from evil of the members of the Church, which is then indirectly signified by the text. Interpreting in terms of *allegoria in factis* was later called 'typological exegesis'.

The assumption of simple *allegoria in verbis* was a way to make sense out of apparent nonsense, and was just one of many strategies used in the interpretation of authoritative texts, secular and sacred alike. After the thirteenth century, a distinction was often made between what a text means *de virtute sermonis* (in virtue of its wording), that is, the meaning the text conveys to anyone knowing the vocabulary and grammar of the language in question, and

what it means *ex intentione auctoris* (in accordance with the intention of the author). By identifying the latter meaning with his own favourite view, a medieval author could avoid declaring his disagreement with an old authority.

In 1340, the Faculty of Arts at the University of Paris forbade its masters to call an authoritative utterance false in virtue of its wording if they thought the author had intended it in a true sense. Instead, teachers were required to say that the text had two senses, one of which was false and the other true, though both senses were 'in virtue of the wording' of the text. According to the reasoning behind the decree, it constitutes no violation of the 'virtue' (force) of the wording if an author intends the words to carry an unusual sense, for words have no semantic 'virtue' independent of their users. The explicit motive for the 1340 prohibition was fear of the consequences of accepting the judgment 'false by virtue of the wording' passed on a Biblical passage, as there might then appear to be a mismatch between the Bible's wording and truth.

8 Multiple signification: systematic use of metaphors (Gilbert and the Porretans)

Gilbert de la Porrée, Bishop of Poitiers from 1142-54 and father of a school of *Porretani*, barely escaped a condemnation for heresy (see Gilbert of Poitiers) for his claim that the proposition *Deus est Deitas* (God is Godhood) is false, while *Deus est Deus Deitate* (God is God in virtue of Godhood) is true. He appeared to be violating the doctrine of divine simplicity (see Simplicity, divine). The Porretans were also ridiculed for tripling the Aristotelian categories. These two pieces of unusual doctrine were in fact part of one theory.

As Augustine and Boethius had made clear, words cannot function the same way when used about creatures and about God. Christian theologians know that 'God is just' is a true proposition because it has the support of Scripture, but they do not fully know what its terms mean. What God is cannot be fully grasped by a human mind, at least not in this life. Nor is it clear what it is for God to be; used about creatures 'is' indicates temporal being, but this cannot be so with the Creator (see God, concepts of). For a creature to be just is to have a certain property, but the doctrine of divine simplicity forbids this interpretation of 'just' in the case of 'God is just'. To speak about God implies transferring terms from the known realm of creatures into the unknown realm of the divine.

Starting from this problem, Gilbert and his followers developed a general theory about transferred language. In their theory, language was originally instituted to deal with the direct results of creation. The terms of this primary language describe the members of natural classes with their substantial and accidental forms. The ten Aristotelian categories are ten classes of words which may be used to describe a subject by signifying a predicate; a predicate in turn is a form that constitutes the subject's being something or in-some-way. In other words, a categorial term signifies a substantial or an accidental form. Only natural things have genuine forms. Hence, properly speaking, the ten categories pertain only to this primary language, that of the *naturalis facultas* (natural domain).

Artifacts, social institutions, values and other consequences of human activity constitute an ontologically secondary group of things to be named, but the *moralis* (behavioural) language used for the purpose is not structurally different from the one used for natural entities. People speak as if houses, prices and moral values were constituted by substantial and accidental properties in the same way as trees and colours. Propositions still have the shape '*x* is *F*', although '*F*' does not signify a genuine form. Most of the words of behavioural language are also loans from the primary language: thus 'high' has been borrowed from the category of quantity to be used in a transferred sense about prices. The semantic contents of the words are of different sorts, but the vocabulary still falls into ten groups, ten quasi-categories, and quite generally the structure of the language remains the same. The behavioural language owes its intelligibility to its retention of the structural features of the natural language.

The second-order language of logic and grammar (*rationalis facultas*) similarly retains the structure of the primary language. There is no real form of equivocation, yet '*x* is equivocal', which has the structure of a genuine predication, may be true; and if it is so, we are justified in talking about the equivocation of *x*.

Finally, there is theological discourse. Words are transferred to new and fundamentally unknown senses. The only way to retain some sort of intelligibility is by obeying the same structural laws, the same syntax, as our primary language. One of the relevant laws might be stated as follows: if '*X*' and '*Y*' are concrete terms, and '*Y*-hood' the abstract term corresponding to '*Y*', then (1) '*X* is *Y*' implies '*X* is *Y* by virtue of *Y*-hood', and (2) '*X* is *Y*-hood' is nonsensical or false. Applied to the case of 'God is… ' the rule yields the result that 'God is God' implies 'God is God by virtue of Godhood', while 'God is Godhood' is an unacceptable proposition. Gilbert did not intend to

undermine the doctrine of divine simplicity. His discovery was that no intelligibility at all is left if one changes syntax and the meaning of the vocabulary at the same time.

9 Multiple signification: metaphor as a key tool in ordinary communication (Roger Bacon)

The mainstream tendency in the thirteenth century was to treat a word's meaning as an unchangeable companion of its sound once the word had been 'imposed' to mean something. This view implied, *inter alia*, that a word does not lose its meaning if such things as it signifies cease to exist; 'donkeys are animals' would be a meaningful (and, most would say, true) proposition even if donkeys had become extinct, and 'donkeys' would signify the same as when donkeys did exist. Similarly, while recognizing a difference between 'the donkey' used to refer to the asinine species and to a particular donkey, most philosophers would deny that a switch from one use to the other amounts to a change of signification.

Around the 1260s, Roger Bacon developed a radically different view. According to Bacon, there is no way in which a word can be imposed simultaneously as the name of an existing and of a non-existing thing, of a universal and of a particular, and so on. Words are instituted and learned with some definite meaning, but it is in every speaker's power to change that meaning. In fact, this is done all the time, and generally tacitly. People are continually confronted with the task of expressing something they have not thought of or talked about before. They habitually solve the problem by means of metaphors, using a term whose meaning is somehow similar to what they now need to express. For example, to someone 'John' may signify a certain animate being. Standing in front of John's dead body, this person may identify the body with the words 'This is John', without intending to say that the corpse is an animate being. Tacitly, 'John' has been given a new meaning (body that used to be the material component of a human being called John). The speaker has performed a new imposition and has rendered the term equivocal.

The correct interpretation of the term in each particular case depends on the speaker's intention. The hearer may not grasp it, but generally the ability of speakers to create metaphors is matched by the ability of listeners to grasp them. To Bacon, metaphor was the individual person's fundamental tool for communicating new thoughts, and we have new thoughts all the time.

10 Multiple signification: the intentionalist current in grammar

Modistic grammar was intended to be an Aristotelian science, with principles of its own. In this the modists had been anticipated by others, but they were particularly insistent on the point that grammar cannot allow the use of information obtainable only from other sources than its own principles. According to the modists, grammar is concerned with grammaticality, not with truth or reality, and a string of words is grammatically well-formed (*congrua*) if it exhibits no mismatched modes of signifying. For this string to be a complete (*perfecta*) sentence, it must fulfil the further requirements of containing a subject (*suppositum*) and a predicate (*appositum*), and of leaving no relations of grammatical dependence between the ingredient words unsaturated. In modism, the lexical content of the sentence's constituent words is grammatically irrelevant, and so are non-linguistic circumstances that might influence the interpretation of an utterance. It follows that *turba ruunt* (the crowd are rushing) is ill-formed, whereas *capa categorica* (a categorical cloak) and *Socrates curret heri* (Socrates will run yesterday) are well-formed expressions.

By contrast, many twelfth-century thinkers had held that such expressions are ill-formed because they are unintelligible. Congruity was required both on the vocal level and on the level of sense. This meant that many putative propositions could be dismissed as not even being grammatically correct sentences. Even lack of referents for an ingredient term could rob an otherwise good sentence of sentencehood and propositionhood.

Constructions such as *turba ruunt* violate rules of concord (in this case, the rule that a verb agrees in number with its subject). However, such constructions could be found in authoritative writers and were generally classed as 'figurative'. For an 'error' to qualify as a figure instead of a mere error, it must have some reasonable excuse: *figura est vitium ratione excusatum*, as the saying went. Moreover, the reasons excusing the deviant use of words were held to be twofold: there are reasons why the deviation is possible, and there are reasons why it is demanded, *ratio qua potest fieri* and *qua oportet fieri*, respectively. Reasons of the first type explain why the deviant expression is intelligible to an audience (for example, *turba ruunt* strongly resembles the correct *turba ruit* and

turba, though of singular number, signifies a plurality of people, so somehow the plurality required by *ruunt* is present to the hearer's mind). Reasons of the second type explain why a speaker would choose the deviant formulation rather than a regular one (for example, the use of a plural verb is needed to give a vivid impression of the many individuals that compose the crowd); that is, such reasons appeal to the speaker's intentions in formulating the message.

Figurative and incomplete constructions invited grammarians to take the situation of communication into account. Further, as in the case of metaphors, appeal could be made to the speakers' intentions as a major factor in deciding the meaning attributable to an utterance. This ran counter to modist principles, but since modism did not have any good tools for dealing with deviant yet intelligible expressions, even modists would sometimes adopt 'intentionalism', as this context-sensitive grammar has been called. Robert Kilwardby and Roger Bacon were among the foremost intentionalists.

In theology, there was an ongoing debate whether the verbal components of the sacraments were in some sense bearers of a power to sanctify along with their power to signify. The less inclined a theologian was to accept this, the more it would be stressed that the sacrament's efficacy depends on the intention of its original institutor and of its administrator (the priest).

11 Multiple signification: primary and secondary signification

Anselm, in his *De grammatico*, introduced a distinction between what a word signifies by itself (*per se*) and what it signifies via something else (*per aliud*). By itself *grammaticus* (literate) signifies literacy plus having (or having literacy); indirectly it signifies a human being to whom having literacy applies. The human being is not directly signified by the word, but anyone who understands it can supply the information that literacy requires a human possessor. Anselm also uses the expression *appellare* as a synonym of *significare per aliud*.

In the twelfth century, it became common to distinguish a word's signification from its appellation or nomination. The things 'called' or 'named' (*appellata*, *nominata*) were whatever things the term applies to considered in their own right, whereas the term was said to 'signify' also or exclusively a concept, a form or the things qua carriers of some form. Thus 'white' might be said to name white things but to signify whiteness. In the thirteenth century, 'nomination' disappeared and 'supposition' replaced 'appellation' as the general term in descriptions of the varieties of reference.

Thirteenth-century writers worked not only with the distinction between signification and supposition, but also with a distinction between primary and secondary (*secund(ari)a* or *ex consequenti*) signification. A concrete accidental term (white) was said to signify primarily a form (whiteness) and secondarily whatever has that form (the thing that is white). Similarly, a substance-term like 'man' might be said to signify primarily human form (humanity) and secondarily the 'aggregate' of matter and form (man). A relative (such as 'parent') could be said to signify one relative thing (the parent) primarily and its correlative (the child) secondarily. Excessive use of the distinction led some to hold that a proposition could be true or false *per se* on account of the primary significates of its terms and have the opposite truth value *per accidens* on account of their secondary significates.

In the next century, William of Ockham used the distinction as a means to keep his ontology lean. He operated with (a) *absolute* terms such as 'dog', which signifies individual dogs and nothing else, and (b) *connotative* terms such as 'white' or 'parent', which directly signify whatever individual things are white or parents but connote (that is, secondarily signify) their whitenesses and children. By treating all terms in categories other than substance and quality as connotative, Ockham could avoid positing special quantitative entities, relative entities and the like.

Buridan made very similar use of the concept of connotation although, confusingly, he stood the twelfth-century terminology on its head, saying that 'white' directly signifies and supposits for something white but connotes or *appellat* its whiteness. Similarly, 'rich' supposits for a man and *appellat* his possessions. Names of fictitious entities (for example, 'chimera') signify nothing directly, but they connote the real entities such monsters are mistakenly supposed to consist of (a snake's tail, a lion's head and so on). Oblique cases of nouns are limited to connoting; thus in 'Plato's donkey is running', Plato is only connoted whereas a donkey is directly signified. Buridan calls what a connotative term supposits its 'matter' and what it connotes its 'form', but the latter does not imply that his *connotata/appellata* are supra-individual things.

12 Multiple signification: properties of terms (supposition, ampliation and restriction)

Handbooks of logic from the later Middle Ages generally devote much space to the so-called 'properties of terms'. The most notable of these are signification, supposition, ampliation, restriction and distribution.

The twelfth-century notion of supposition may have part of its background in Trinitarian theology, which distinguished between *suppositio personalis* and *suppositio essentialis* according to whether a term stands for one of the divine persons or the divine essence. However, the further development of supposition theory took place in the context of logic. The core idea of the theory is that a term does not always represent the same objects, even though it retains its meaning. In different contexts, the term refers to different subsets of the totality of objects of which it may be truly predicated. The proposition 'all the doctors will dispute next year' is ambiguous as to whether it claims that all presently existing doctors will dispute next year or that all the doctors existing next year will be disputing then. This was expressed in the rule that a general term acting as subject for a verb of future tense supposits for things present or future. Sometimes this rule comes in the form 'things present and future'. The first formulation shows an understanding of the term's *supposita* as the minimal set of objects that will verify the proposition; the second formulation takes the term's *supposita* to be the conjunction of those sets.

Another core idea was that a term's *supposita* are the individuals falling under it in some particular sentential context, but from the very beginning this idea was somewhat obfuscated by attempts to allow the notion of supposition also to cover other variations in the way a term may be understood. A distinction between personal and material supposition was commonly accepted. In personal supposition, the word 'donkey' stands for members of the asinine species; in material supposition, 'donkey' stands 'for itself', that is, for the word 'donkey', as in 'donkey is a noun'. Counting material use as a sort of supposition was a complication of the theory, and it was often unclear what exactly was meant by 'itself' in the formula 'standing for itself'. Another complication was the acceptance in the thirteenth century of a 'natural supposition': lifted out of any sentential context, a term was said to have natural supposition for the totality of objects falling under its signification. This could only make sense in an age when signification was not taken as a relation between a word and extramental particulars but between a word and some non-particular entity. Many authors further accepted a *suppositio pro significato* or *suppositio simplex*, claiming that in 'the donkey is a species', 'donkey' stands for the universal signified by 'donkey'.

The notion of supposition was most successful as a tool for dealing with tense, quantification and the effects of intentional verbs (such as 'know' and 'promise'). In the fourteenth century, Ockham and Buridan removed many of the inconsistencies in the knowledge inherited from the previous century. The notion of supposition was inextricably linked to those of ampliation (widening) and restriction. The idea is this: the set of objects for which a term would otherwise stand can be made bigger or smaller by the occurrence of some other word. Typically the neutral situation was taken to be that of a subject of a present-tense proposition containing neither modal nor intentional verbs - for example, 'every man is running' - in which 'man' was said to supposit for presently existing men. The addition of an adjective ('every young man is running') will restrict the set of verifiers by excluding old men. Saying 'may be running' instead of 'is running' extends the range of the term by adding an alternative set of verifiers, that of men who are yet to be (according to some authors) or that of possible men (according to others).

Fourteenth-century logic usually operated with four domains (sometimes called 'tenses'): the past, the present, the future and the possible. Marsilius of Inghen proposed to add a fifth 'tense' or domain, that of the imaginable, in order to provide verifiers for propositions of the type '*X* is imagined', in which '*X*' refers to an object that is indeed imaginable even though it may not be physically possible. According to his proposal, the verb 'is imagined' ampliates its subject so that it supposits not only for actually existing things but also for imaginable ones. This was a considerable step in the direction of developing a theory of the language of fiction, a subject to which scholastic philosophers had paid little attention save for one much-discussed example: propositions about chimeras. Chimeras were supposed to be impossible entities, though it was often unclear what sort of impossibility was involved. Marsilius' master, John Buridan, had sided clearly with logical impossibility, and to underline this he had ended up changing the medieval standard definition of a chimera, 'animal composed of lion, cow and young woman', into 'animal composed of parts that cannot possibly be put together'. However this invited the question of what to do about logically possible, and hence imaginable, things that nature does not allow. This was the background of Marsilius' proposal of an 'imaginable tense'.

13 Multiple signification: ambiguity

After about 1150, Aristotle's *Sophistici Elenchi* provided the general framework for most discussions of ambiguity. Aristotle used six fallacies 'in speech'. Following Galen, medieval writers usually took these six fallacies to be due to three types of ambiguity (*duplicitas*), occurring in two types of expressions, single words (*dictiones*) or complexes of words (*orationes*). The resulting system is described in Table 1:

Table 1

Expression	Duplicity		
	actual	potential	imaginary (*fantastica*)
Simple	equivocation	accent	figure of speech
Complex	amphiboly	composition and division	

An actual ambiguity was supposed to be present any time the expression is uttered. The standard example of equivocation was *canis*, which can refer to (1) a dog, (2) a dogfish, or (3) the Dog Star. According to late thirteenth-century doctrine, such a term could be disambiguated only by the immediate addition of a determination suitable to only one of the significates. Thus 'dog' was supposed to have only sense (1) in 'a barking dog' while keeping all three senses in 'a dog is barking'. Aristotelian equivocation was designed primarily for nouns, but medieval writers sometimes used equivocation in all parts of speech, even pronouns and prepositions. For example, some held that 'this' is equivocal with respect to all things that may be pointed at, and the preposition *de* was considered equivocal because it can indicate local origin or material composition, as in *Socrates est de terra*, which can mean (1) 'Socrates is *from* a country' or (2) 'Socrates is *composed of* earth'.

In genuine equivocation, one sense cannot be derived from another. Each sense derives from an independent imposition when a sound was chosen, without concern for whether it had already been selected to label some other thing. According to the principles of modist grammar, this would mean that there could not strictly be such a thing as an equivocal noun. There could be one phonetic entity common to three significates, but then there would be as many nouns as there are significates. When this was realized in the 1270s, an attempt was made to save the essentials of the theory while preserving the intuition that *canis* is exactly one noun with the same grammatical properties no matter which sense is intended by the speaker or needed to make the proposition true in which the word occurs. It was proposed that there are both active and passive modes of signifying, with the active ones residing in the word and their passive counterparts (indistinguishable from modes of being) in the things signified. One active mode may correspond to three different passive ones, it was claimed, and syntactic properties depend on the active modes. This was not a happy solution, since it introduced two levels of grammatical structure with the surface structure determining syntax. As so often in the history of linguistics, ambiguity had proved a hard test for a theory (see Ambiguity).

In the principal type of equivocation, the different senses of the word are not ordered. However, standard doctrine also accepted types of equivocation in which one sense could be described as primary or proper and the other(s) as secondary, improper or transferred, as when 'running' primarily describes fast foot-propulsion, but metaphorically describes the movement of a river. Thirteenth-century writers adopted the name 'analogy' for equivocation in which one thing is signified primarily and one or more others secondarily. If a term signifies all its significates, equally it always needs a determination in order to be disambiguated. An analogical term unaccompanied by any determination was held to represent only its primary significate. An Aristotelian example was often used: 'healthy' primarily signifies a property of animate beings, while secondarily it signifies (a) a property of animal excretions, namely that of indicating that the animal they come from enjoys good health, and (b) a property of food

and drugs, that of being conducive to the health of some animate being. To actually signify (a) or (b), 'healthy' needs to be joined with words like 'urine' or 'potion'. This particular type of analogy was called *analogia attributionis*. A considerable debate arose concerning another example, 'man', which some held to signify primarily a compound of body and rational soul and secondarily a mere body. In this view, the primary significate is represented by 'man' in 'this is a man' and the secondary one by the 'man' of 'this is a dead man', and so the latter proposition does not entail the former.

Some modists assumed that the significative function (*ratio significandi*) assigned to an analogical term on its imposition was a complex one selecting different significates in different contexts. Radulphus Brito gave up considering the multiple meaning a property of the word itself prior to its actual use. The secondary signification comes about, he claimed, when a speaker uses a word improperly and a listener catches his intention thanks to a charitable interpretation of an utterance that would make little sense if taken literally. In so claiming, however, Brito sacrificed the modist axiom that only two factors contribute to the sense of an utterance: lexical (imposition-given) meaning and modes of signifying.

In the thirteenth century, it became standard doctrine that *ens* (being) is analogical as to substance and accidents. Attempts were made to use this distinction as a model for the relation between the use of terms for both creatures and the creator, and a long debate about the analogy or univocity of being ensued (see Being).

The notion of potential ambiguity led to a theory according to which the *litterae* (letters, phonemes) are the matter of an expression, which can be actualized only by the addition of a form, the 'way of pronouncing' (*modus pronuntiandi*), consisting in stress, pauses and the like. For example, in English the letters 'i-n-v-a-l-i-d' are the common matter of two distinct actual words, '*in*valid' and 'inv*al*id'. The fallacy of composition and division was used, *inter alia*, to explain cases in which a string of words can be interpreted either as an atomic or as a molecular proposition; thus *falsum est Socratem esse hominem si est animal* can be read as an atomic proposition ('that-Socrates-is-a-man-if-he-is-an-animal is false') or a molecular proposition ('that-Socrates-is-a-man is false, if he (or: there) is an animal'). The distinction between a 'composite' and a 'divided' sense also played a central role in the analysis of modal propositions. Often, *possibile est sedentem stare* was said to mean 'that-a-man-stands-while-sitting is possible' in the composite (or *de dicto*) sense, and 'a sitting man has the possibility of standing' in the divided (or *de re*) sense.

The notion of imaginary ambiguity was always somewhat unclear, but 'figure of speech' became a head under which the scholastics could examine the relation between (onto)logical and grammatical categories, for example, action versus active voice, substance versus substantive and individual versus singular number.

Use-mention ambiguity was usually dealt with in terms of supposition. In 'man is an animal', 'man' supposits personally (that is, for things signified by the word 'man'), while in 'man is monosyllabic', 'man' supposits materially (for phonetic entities equiform with the word pronounced). On the other hand, ambiguous supposition was often classified as a mode of either equivocation or figure of speech.

Logicians classified words as either categorematic (roughly = nouns + verbs) or syncategorematic (mostly quantifiers, modal terms, prepositions and conjunctions). There were few ancient precedents for inquiry into the contribution of syncategoremata to the sense of a proposition, but a major research effort was made in the later Middle Ages, especially with regard to ambiguity traceable to the occurrence of such logical operators, including scope ambiguity.

14 The significate of propositions

Any syntagm could be called an *oratio*, whether a sentence fragment such as 'a white horse' or a whole sentence (see Propositions, sentences and statements). Among the different types of sentences, the declarative was both the grammarians' preferred and the logicians' sole object of study. In logic, the technical term was *propositio*, and the standard definition (from Boethius) was *propositio est oratio verum vel falsum significans* (a proposition is a syntagm that signifies a truth (or some true thing) or a falsehood (or some false thing)). Propositions were said to be either categorical (that is, atomic) or hypothetical (that is, molecular; in medieval terminology, conditionals are only one among several types of hypothetical propositions). The Boethian definition suggests that there are entities - truths and falsehoods - signified by propositions. In the twelfth century, the significate of a proposition was usually called *enuntiabile* (statable) or *dictum (propositionis)*, (what is said (by the proposition)). Thus 'Socrates is

running' (*Socrates currit*) signifies the *dictum* that-Socrates-is-running (*Socratem currere*), and the string of words 'that Socrates is running' is the name of the *dictum* (*nomen dicti*). It was commonly, though not universally, assumed that the objects of knowledge and belief are *dicta*.

The ontological status of *dicta* was controversial. Nominalists could not give them full reality, for that would be a repetition of the error of hypostatizing the significates of single words. On the other hand, a *dictum* is either a truth or a falsehood, and some truths at least might seem to be eternal; but if *dicta* were eternal, there would be other eternal things than God, which was scarcely an acceptable thought. The majority view seems to have been that *dicta* are a sort of quasi-thing.

Controversy also surrounded the nominalist thesis *quod semel est verum semper est verum* (what is true at one time is always true). This thesis was theologically motivated. The pre-Christian patriarchs were assumed to have believed that Christ would be born, and that was a truth; Christians believe that Christ has been born, and that is a truth. Do Christians believe in the same truth as did the patriarchs? Or did the truth in which the patriarchs believed cease to be a truth when Christ was born? In the nominalist view, 'Christ will be born' and 'Christ has been born' express the same truth supposing that they are pronounced before and after his birth, respectively. The same true *dictum* has two differently tensed names, 'that Christ will be born' and 'that Christ has been born'. The nominalist thesis amounts to a claim that the class of linguistic expressions that are verified by the same event or events express the same *dictum*. Non-nominalists replied that the differently tensed 'that' clauses are or signify different statables; some tried to escape unfortunate theological consequences by denying that statables are what knowledge and belief have for their objects.

After the twelfth century, interest in statables abated somewhat, but it came back strongly in the fourteenth. Most authors then shared the belief that propositions, whether mental or spoken, are individual entities existing in time. Previously there had been a tendency to neglect the difference between proposition token and proposition type, but now the distinction came into the foreground, especially thanks to John Buridan. He also contributed to the popularity of a three-part analysis of propositions into subject, copula and predicate, with the copula treated as a term in its own right. The copula of a mental proposition could be described as a 'complexive concept' uniting the extremes. Pierre d'Ailly, while accepting the three-part analysis, insisted that ontologically a mental proposition is a unity; it is tripartite only in the sense that it is a cognition equivalent to the conjunction of three other cognitions.

Philosophers influenced by Ockham often held that the objects of knowledge and belief are mental propositions. According to Ockham, a proposition's truth is nothing other than the proposition itself, and it has no other significate than the individual things to which its terms refer. Some found this implausible, *inter alia* because truth would then seem to depend on the existence of propositions in human minds rather than on external reality. Instead, they (most famously Gregory of Rimini), took the object of knowledge and bearer of truth to be a state of affairs, *complexe significabile (tantum)* (something that can (only) be signified in a complex way, for example by a proposition, not by a term). The same quasi-entity could also be called *totale* (or *adaequatum*) *significatum propositionis* (the total (or: adequate) significate of a proposition), as opposed to the significates of each of its ingredient terms. Gregory vigorously denied that total significates are existing entities, whether mental or extramental. Consequently, he thought, God's uniqueness as an eternal being is not compromised by claiming that even before the beginning of the world 'complexly signifiables' were true or not true in virtue of the sign and source of all truth, God himself.

Walter Burley held the unusual view that just as vocal terms have real things as their ultimate significata, so vocal propositions ultimately signify propositions in external reality. However, he seems never to have reached a satisfactory answer to the question of what sort of thing such a *propositio in re* is. In one of his formulations, a real proposition has the things signified by its subject and predicate for its matter and a composition or division of that matter (in affirmation or negation) contributed by the mind for its form.

15 Syntax

The explicit aim of all medieval grammar was to spell out the rules of grammaticality (*congruitas*), or in other words, syntactical rules. The key term in grammatical syntax was *constructio*, standardly defined as *congrua dictionum ordinatio* (congruous linking of words), and most attention was paid to the pairwise linking of words

rather than to more complex linguistic entities. The elementary rules of concord were easily describable in terms of grammatical accidence; thus substantive nouns and adjectives were both said to have the accidents of gender, number and case, and the rule 'a substantive and its adjective must agree in gender, number and case' was thus a rule on how to match the accidents of one word with those of another. In modistic grammar, such concord could be described as a 'proportionality of modes of signifying', with the idea being the same, that certain secondary semantic features of one of the words to be construed require or forbid certain semantic features of the other word.

The relationship between the words to be construed was most often described in terms of determination, government (*regimen*), demand (*exigentia*), or dependence with the last term dominating in modistic grammar. In about 1300, a construction was taken to be the union of a *dependens* (dependent, unsaturated constituent) and a *dependentiam terminans* (a constituent that saturates the dependence). Just as in logic a proposition is a union or composition of subject and predicate, most simply expressed in the form 'noun + verb', so in grammar a sentence is a union of *suppositum* and *appositum*, most simply expressed as 'noun + verb', with the verb being the dependent constituent requiring saturation such as can be provided by the noun that signifies *per modum per se stantis* (purports that its significate is a self-sufficient entity). In the complex sentence 'Socrates goes to the church' the verb 'goes' has two relations of dependence, one being saturated by 'Socrates', another by '(to) the church'. As far as the verbs are concerned, the medieval notion of dependence had approximately the same function as the modern notion of valency.

16 Informative and performative locutions

While most medieval linguistic thought concentrated on descriptive propositions, a conceptual apparatus for talking about performative locutions did exist, at least since the early thirteenth century. This was provided by a distinction between *actus significatus* (or *conceptus*) and *actus exercitus*, act(ion) signified or conceived versus act(ion) performed. The distinction was used in several ways.

First, if *A* says 'Bravo!' when hearing *B* say 'They sacrificed themselves', the exclamation may either comment on the act of self-sacrifice signified by *B*'s statement or on the act performed by *B* in uttering the statement. This raises the question of whether in the first case 'Bravo' is construed as an adverb with the word 'sacrificed' but in the second case with an event, not with a word.

Second, some words signify what others perform. The words 'I address you' signify an action that is performed by saying 'Hello'; 'or' and 'every' perform - or are means of performing - what 'disjoins' and 'distributes' signify. Some logicians held that syncategoremata (that is, logical operators) were purely performative, having no signification besides their function (*officium*). According to Ockham, '*B* is predicated of *A*' should not be confused with '*A* is *B*': in the former proposition predication is signified, in the latter it is performed.

Third, in some situations words which are usually informative (signify an action) become performative: when a priest says 'I baptize you', his words are instrumental in performing what they describe. Finally, some thought a confusion of signification with performance was the source of the Liar paradox, the act of saying signified by 'I am saying a falsehood' being confused with the act of saying performed by uttering those words.

17 The diversity of languages

Medieval theoreticians were aware of the possibility of non-vocal sign systems fulfilling the same role as ordinary languages, but spent little time on the investigation of any other non-vocal communication than that between angels, who were supposed to use no material signs because of their immateriality. The theoreticians were, of course, aware of the diversity of human languages but usually merely acknowledged the phenomenon, noting that different peoples can use different conventional signs for the same things. The modists of the late thirteenth century produced an account of how the shared human rationality and shared external reality resulted in one common linguistic structure (grammar), capable of accidental diversification through the arbitrary choice of material (sounds) to carry the sign function. Modists did not typically raise the question of a possible common origin of the different languages' phonetic manifestations. Medieval writers who did raise the question usually agreed that the first human language, that of Adam, was Hebrew, and that there was some sort of connection linking Hebrew to the many languages spoken after the Babylonian confusion, but few apart from Roger Bacon and Dante had any clear ideas of linguistic development.

Dante Alighieri presented a remarkably well-developed theory of linguistic change in his *De vulgari eloquentia*. Dante operates with two types of language: (a) such as is the maternal tongue of certain people, (b) such as is no-one's primary language. Languages of type (a), the vernaculars, are subject to change, while type (b) languages, the 'grammatical' ones such as Latin and classical Greek, are exempt from change, their form having been determined once and for all by codified rules. The 'grammatical' languages are artificial; they have been created to avoid the breakdown of communication that spatial and temporal distances cause in natural languages. Contrary to the humanists of later periods, Dante did not see Latin as Cicero's maternal tongue, though he was clearly aware that it was based on ancient Roman vernacular.

To Dante the natural languages, the vernaculars, are the nobler ones. All derive from the language that Adam acquired on acquiring his soul. Thus far it might appear as if Dante followed modist teaching, for if Adam's acquisition of language was inseparable from his acquisition of soul, language is innate in humans. However, whereas the modists only claimed that all languages share an abstract structure, Dante saw them all as historical developments of one original language, that of Adam, which was not just an abstract structure but a fully fledged language with phonetics, vocabulary and syntax. He tried to make sense of the Babylonian confusion by taking it to represent the emergence of sociolects, each craft developing its own way of speaking. However, his finest observations concern the way language changes with time and how geographical isolation can make two branches of a language develop in different ways. Applying his theory to European languages, he assumed that all derive from three languages that must have arrived in very ancient times. One of these was the ancestor of Northern and Eastern European languages (about which Dante obviously knew little; Roger Bacon had been better informed), the second of the Romance languages, and the last of Greek.

Since Dante's purpose was to persuade his readers that Italian could serve as a language of refined literature, he had to face the argument that is always brought up when a new literary language is being formed, namely that there is no such common language with a describable form of its own but only a large number of local dialects. His counterclaim was that it makes sense to have a supra-regional Italian language, identical with no particular dialect but acceptable to educated people everywhere.

18 An artificial language

No medieval author envisaged an artificial, uninterpreted language. Nevertheless, some developments in logic pointed in that direction. The use of variables was known from Aristotelian syllogistics and was employed by medieval writers in the same way as Aristotle had done. Moreover, a number of standard examples developed into quasi-variables, the name 'Socrates' having any human proper name for its value, 'Caesar' representing any proper name of a human being from the past, 'currit' any present-tensed active verb, and so on. Thus, if a logician asked whether the proposition 'Caesar is a man' is true, he was asking about propositions with the proper name of a no-longer-existing individual for their subject term. Buridan explicitly recognized that the standard examples in jurisprudence, 'Titius' and 'Bertha', were not genuine proper names; they do not elicit singular concepts, but common ones.

Logicians also came close to creating an artificial language when they stated rules of supposition, often reflecting tendencies rather than hard and fast rules of Latin. Thus according to many logicians, *asinus cuiuslibet hominis currit* equals 'there is a donkey such that it belongs to just any man, and this donkey is running' while *cuiuslibet hominis asinus currit* means 'for just any man it is true that his donkey runs'. Though not quite aware of the fact, these logicians had created fragments of an artificial disambiguated language. This was to earn them much scorn from humanists, who could see no good in unnatural Latin.

See also: Aristotelianism, medieval; Language, philosophy of; Language, Renaissance philosophy of; Logic, medieval; Natural philosophy, medieval §8; Nicholas of Autrecourt; Nominalism; Semantics; William of Ockham §6

STEN EBBESEN

References and further reading

Biard, J. (1989) *Logique et théorie du signe au XIVᵉ siécle*, Paris: Vrin.(Especially valuable as an introduction to some lesser-known thinkers from the fourteenth century.)

Bursill-Hall, G.L., Ebbesen, S. and Koerner, K. (eds) (1990) *De ortu Grammaticae: Studies in medieval*

grammar and linguistic theory in memory of Jan Pinborg, Studies in the History of the Language Sciences 43, Amsterdam: Benjamins.(Articles on medieval logic and grammar.)

Ebbesen, S. (1987) 'The Way Fallacies Were Treated in Scholastic Logic', *Université de Copenhague. Cahiers de l'Institut du Moyen Age Grec et Latin* 55: 107-34.(Survey article with extensive bibliography.)

Ebbesen, S. (ed.) (1995) *Sprachtheorien in Spätantike und Mittelalter*, Geschichte der Sprachtheorie 3, Tübingen: Narr.(Articles in English, German and French on a broad variety of themes. Also useful as a bibliographical guide.)

Kneepkens, C.H. (1987) *Het iudicium constructionis. Het leerstuk van de constructio in de 2de helft van de 12de eeuw*, 4 vols, Nijmegen: Ingenium.(On twelfth-century grammarians' theories of syntax. Discussion in vol. I followed by text editions in vols II-IV.)

Kneepkens, C.H. (1990) 'On Medieval Syntactic Thought with Special Reference to the Notion of Construction', *Histoire Épistémologie Langage* 12: 139-76.(Can be used as a guide to the scarce literature on medieval syntactic theory.)

Kretzmann, N. (ed.) (1988) *Meaning and Inference in Medieval Philosophy: Studies in Memory of Jan Pinborg*, Synthese Historical Library 32, Dordrecht: Kluwer.(Contains several relevant papers.)

Kretzmann, N., Kenny, A. and Pinborg, J. (eds.) (1982) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press.(Several relevant articles.)

Libera, A. de (1996) *La querelle des universaux. De Platon à la fin du Moyen Age*, Paris: Éditions du Seuil. (Provides an up-to-date account of medieval theories of universals.)

Marmo, C. (1994) *Semiotica e linguaggio nella scolastica: Parigi, Bologna, Erfurt 1270-1330*, Roma: Istituto Storico Italiano per il Medio Evo.(A thorough study of modistic theory; uses both grammatical and logical sources.)

Nuchelmans, G. (1973) *Theories of the Proposition: Ancient and Medieval Conceptions of the Bearers of Truth and Falsity*, North-Holland Linguistic Series 8, Amsterdam: North Holland.(A classic with strong impact on later research.)

Panaccio, C. (1991) *Les mots, les concepts et les choses*, Montreal, Que.: Bellarmin, and Paris: Vrin.(A penetrating analysis of Ockham's theory of language.)

Pinborg, J. (1967) *Die Entwicklung der Sprachtheorie im Mittelalter*, Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters, Texte und Untersuchungen 42.2, Münster: Aschendorff, and Copenhagen: Frost-Hansen.(The classic on modistic grammar.)

Pinborg, J. (1972) *Logik und Semantik im Mittelalter. Ein Überblick*, Stuttgart: Frommann-Holzboog.(Survey with extensive quotations of original texts.)

Pinborg, J. (ed.) (1976) *The Logic of John Buridan*, Copenhagen: Museum Tusculanum.(Somewhat outdated, but still useful as an introduction to Buridan's semantics.)

Pinborg, J. (1984) *Medieval Semantics, Selected Studies on Medieval Logic and Grammar*, ed. S. Ebbesen, London: Variorum.(Several relevant papers.)

Rijk, L.M. de (1962-7) *Logica Modernorum*, 2 vols in 3 parts, Assen: Van Gorcum.(The richest single collection of twelfth-century logic texts with extensive discussion by the editor; partly outdated, but still invaluable.)

Rijk, L.M. de (1989) *Through Language to Reality: Studies in Medieval Semantics and Metaphysics*, ed. E.P. Bos, Northampton: Variorum.(Several relevant papers.)

Rijk, L.M. de and Braakhuis, H.A.G. (eds.) (1987) *Logos and Pragma: Essays on the Philosophy of Language in Honour of Professor Gabriel Nuchelmans*, Nijmegen: Ingenium.(Several papers about the meanings of terms and propositions.)

Rosier, I. (1983) *La Grammaire spéculative des Modistes*, Lille: Presses Universitaires de Lille.(Valuable survey of modist doctrine.)

Rosier, I. (1994) *La parole comme acte. Sur la grammaire et la sémantique au XIIIe siècle*, Paris: Vrin.(On the 'intentionalist' current in grammar and on a wide range of linguistic topics in medieval grammar, philosophy and theology, with special attention paid to Roger Bacon; rich bibliography.)

Vineis, E. and Maierù, A. (1990) 'La linguistica medioevale', in G.C. Lepschy (ed.) *Storia della linguistica* II, Bologna: Il Mulino, 11-168.(A good introduction to the grammatical tradition in particular, with rich bibliography.)

Language, philosophy of

Philosophical interest in language, while ancient and enduring (see Language, ancient philosophy of; Language, medieval theories of; Language, Renaissance philosophy of; Language, early modern philosophy of), has blossomed anew in the past century. There are three key historical sources of the current interest, and three intellectual concerns which sustain it.

Philosophers nowadays often aspire to systematic and even mathematically rigorous accounts of language; these philosophers are in one way or another heirs to Gottlob Frege, Bertrand Russell, Ludwig Wittgenstein and the logical positivists, who strove to employ rigorous accounts of logic and of meaning in attempts to penetrate, and in some cases to dispel, traditional philosophical questions (see Logical positivism). Contemporary philosophers, too, are often attentive to the roles that philosophically interesting words (like 'know', 'true', 'good' and 'free') play in ordinary linguistic usage; these philosophers inherit from 'ordinary language philosophers', including G.E. Moore, J.L. Austin and again Wittgenstein, the strategy of finding clues to deep philosophical questions through scrutiny of the workaday usage of the words in which the philosophical questions are framed (see Ordinary language philosophy).

Philosophical interest in language is maintained by foundational and conceptual questions in linguistics, quintessentially philosophical problems about the connections between mind, language and the world, and issues about philosophical methodology. These springs sustain a rich and fascinating field of philosophy concerned with representation, communication, meaning and truth.

1 Philosophy of linguistics

Language is an impressive and fascinating human capacity, and human languages are strikingly powerful and complex systems. The science of this capacity and of these systems is linguistics. Like other sciences, and perhaps to an unusual degree, linguistics confronts difficult foundational, methodological and conceptual issues.

When studying a human language, linguists seek systematic explanations of its *syntax* (the organization of the language's properly constructed expressions, such as phrases and sentences; see Syntax), its *semantics* (the ways expressions exhibit and contribute to meaning; see Semantics), and its *pragmatics* (the practices of communication in which the expressions find use; see Pragmatics).

The study of syntax has been guided since the 1960s by the work of Noam Chomsky, who, in reaction to earlier behaviourist and structuralist movements in linguistics (see Behaviourism, analytic; Behaviourism, methodological and scientific; Structuralism in linguistics; Saussure, F. de), takes an unapologetically cognitivist approach. Human linguistic capacities, he holds, issue from a dedicated cognitive faculty whose structure is the proper topic of linguistics. Indeed, Chomsky construes at least the study of syntax and (large parts of) semantics as attempts to uncover cognitive structures. Finding impressive commonalties among all known natural languages, and noting the paucity of evidence and instruction available to children learning a language, Chomsky suggests that surprisingly many features of natural languages stem from innate characteristics of the language faculty (see Chomsky, N.; Language, innateness of).

Whereas contemporary philosophers have tended to stay at a remove from work in syntax, discussing rather than doing it, semantics is another matter entirely. Here many of the great strides have been made by philosophers, including Gottlob Frege, Bertrand Russell, Ludwig Wittgenstein, Rudolf Carnap, Richard Montague and Saul Kripke. (However, quite a number of linguists and logicians who do not call themselves philosophers also have contributed heavily to semantics.) One major strand in semantics in the past century has consisted in the development and careful application of formal, mathematical models for characterizing linguistic form and meaning (see Semantics, game-theoretic; Semantics, possible worlds; Semantics, situation).

Pragmatics, at least as much as semantics, has benefited from the contributions of philosophers. Philosophical interest in pragmatics typically has had its source in a prior interest in semantics - in a desire to understand how meaning and truth are situated in the concrete practices of linguistic communication. The later Wittgenstein, for instance, reminds us of the vast variety of uses in which linguistic expressions participate, and warns of the danger of assuming that there is something aptly called their *meanings* which we might uncover through philosophy. J.L.

Austin seeks in subtleties of usage clues to the meanings of philosophically interesting terms like 'intentional' and 'true'. Austin keeps a careful eye to the several different things one does all at once when one performs a 'speech act' (for instance: uttering a sound, voicing the sentence 'J'ai faim', saying that one is hungry, hinting that one's companion might share their meal, and causing them to do so). His taxonomy has provided the basis of much subsequent work (see Speech acts; Performatives). H.P. Grice, while critical of some of Austin's methods, shared the aim of distilling meaning from the murky waters of use. Grice portrays conversation as a rational, cooperative enterprise, and in his account a number of conceptions of meaning figure as central strategies and tools for achieving communicative purposes. Grice's main concern was philosophical methodology (see §3), but his proposals have proven extremely popular among linguists interested in pragmatics (see Communication and intention; Meaning and communication). Recently, philosophers and linguists have become increasingly persuaded that pragmatic concerns, far from being mere addenda to semantics, are crucial to the questions of where meaning comes from, in what it consists, and how the many incompletenesses and flexibilities in linguistic meaning are overcome and exploited in fixing what speakers mean by their words on particular occasions (see Pragmatics; Implicature; Metaphor).

Our focus on language should not omit a field of study with a rather broader scope, namely *semiotics*, which is the study of signs and signification in general, whether linguistic or not. In the view of the scholars in this field, the study of linguistic meaning should be situated in a more general project which encompasses gestural communication, artistic expression, animal signalling, and other varieties of information transfer (see Semiotics; Animal language and thought).

2 Meaning: language, mind and world

Philosophy aims at intellectually responsible accounts of the most basic and general aspects of reality. Part of what it is to provide an intellectually responsible account, clearly, is for us to make sense of our own place in reality - as, among other things, beings who conceive and formulate descriptions and explanations of it.

In framing issues about our roles as describers and explainers, philosophers commonly draw a triangle in which lines connect 'Language', 'Mind' and 'World'. The three lines represent relations that are keys to understanding our place in reality. These relations in one or another way constitute the *meaningfulness* of language.

Mind ↔ *World.* Between Mind and World there are a number of crucial relations studied by philosophers of mind. Among these are perception, action, the mind's bodily constitution and intentionality (the mind's ability to think *about* what is in the world) (see Mind, philosophy of).

Mind → *Language.* Using and understanding language is a heavily mental activity. Further, this activity seems to be what the real existence of meaningful language consists in. In short, mind invests *meaning* in language.

Theorists of language focus on the Mind/Language connection when they consider *understanding* to be the cornerstone concept, holding, for instance, that an account of meaning for a given language is simply an account of what constitutes the ability to understand it (see Meaning and understanding). Philosophy has seen a variety of accounts of wherein understanding consists. Many have been attracted to the view that understanding is a matter of associating the correct ideas or concepts with words (see, for instance, Locke, J.; Frege, G.; Language of thought). Others have equated understanding with knowing the requirements for accurate or apt use of words and sentences (see, for instance, Davidson, D.; Dummett, M.A.E.). Still others find the key to understanding in one's ability to discern the communicative goals of speakers and writers (see, for instance, Grice, H.P.), or more directly in one's ability to 'pass' linguistically, without censure (see, for instance, Wittgenstein, L.). Certainly, these approaches do not exclude one another.

Some philosophers focus more on production than consumption - on the speaker's side of things - analysing linguistic meaning in terms of the goals and practices of speakers, and in terms of relations among communities of speakers (see Grice, H.P.; Communication and intention; Language, conventionality of; Language, social nature of).

Many of the philosophers who see understanding and use as the keys to linguistic meaning have held that the meaningfulness of language in some sense derives from mental content, perhaps including the contents of beliefs, thoughts and concepts. This enhances the interest of *cognitive semantics*, which is a thriving field of study (see

Semantics; Semantics, conceptual role; Semantics, informational; Semantics, teleological; Concepts).

It has not gone unquestioned that mind indeed can assign meaning to language, and in fact scepticism about this has figured quite prominently in philosophical discussions of language. Wittgenstein has been read as at least flirting with scepticism that there is anything our minds *can* do that would constitute meaning one thing rather than another (see Wittgenstein, L. §§10-12; Meaning and rule-following; Private states and language). W.V. Quine, starting from the thought that meaning is whatever good translation captures, and on arguments that good translation is not squarely dictated by any real facts, concludes that meaning is highly indeterminate. Quine is not alone in the view that linguistic and mental meaning are best seen not as 'out there' to be discovered, but rather as partly constituted or constructed by our practices of interpreting and translating (see Quine, W.V.; Davidson, D.; Dennett, D.C.; Lewis, D.K.; Radical translation and radical interpretation).

Language → Mind. If mind assigns meaning to language, so also language *enables* and *channels* mind. Acquiring and trafficking in a language brings one concepts, thoughts and habits of thought, with all sorts of consequences (see Sapir-Whorf Hypothesis; Linguistic discrimination; Language and gender). Indeed, having language is so crucial to our ability to frame the sophisticated thoughts that appear essential to language-use and understanding, that many doubt whether mind is 'prior' to language in any interesting sense (see Meaning and communication; Davidson, D.).

Language ↔ World. Since language is the vehicle of our descriptions and explanations of reality, philosophers are concerned about what if anything makes for a *true* or *apt* characterization of reality. Philosophers have these concerns for reasons of philosophical methodology (which we will come to in a moment), but also owing to the naturalness and plausibility of a certain picture of meaning.

According to this picture, the key to meaning is the notion of a *truth-condition*. A statement's meaning determines a condition that must be met if it is to be true. For example, my statement 'Ireland is larger than Manhattan', given what it means, is true just in case a certain state of affairs obtains (namely, a certain island's being larger than a certain other island). According to the truth-conditional picture of meaning, the core of what a statement *means* is its truth-condition - which helps determine the way reality is *said to be* in it - and the core of what a word means is the contribution it makes to this (perhaps, in the case of certain sorts of word, this would be what the word *refers* to) (see Semantics; Meaning and truth; Reference).

While the truth-conditional picture of meaning has dominated semantics, a serious challenge has been presented by philosophers, including Michael Dummett, who urge that the key to meaning is a notion of *correct use*. According to this alternative picture, the core of a sentence's meaning is the rule for its appropriate utterance. Of course, the two pictures converge if sentences are correctly used exactly when they are true. The interest of the distinction emerges only when (a 'realist' conception of) truth is dislodged from this role, whether because of scepticism about truth itself, or because truth is seen as too remote from the crucible of social practice to be the meaning-relevant criterion for correct use (see Realism and antirealism; Intuitionistic logic and antirealism; Meaning and verification; Dummett, M.A.E.; Truth, pragmatic theory of; Truth, deflationary theories of; Truth, coherence theory of; Truth, correspondence theory of). The challenge illustrates a sense in which the Mind/Language and Language/World connections can seem to place a tension on the notion of meaning (meaning is whatever we cognitively *grasp*, but the meaning of language just is its bearing on the world).

3 Linguistic philosophy

Apart from language's interest as a target of science and its centrality to our self-conception as describers of reality, language plays a key methodological role in philosophy. It is this role perhaps more than anything else that has explained the continued close attention paid to language in the past century by philosophers working in such varied areas as epistemology, aesthetics, ethics, metaphysics, the philosophy of science and the philosophy of mind.

The methodological role of language in philosophy is most easily explained by example. A philosopher is interested in the nature of value; they want to know *what goodness is*. Language enters when they observe that goodness is what is attributed when we *say* of a thing that it 'is good'. So the philosopher focuses on certain statements, and seeks an understanding of what such statements *mean* and in general of how they work. They explore whether such statements are ever objectively true or false, whether their truth or aptness varies from

speaker to speaker, whether a satisfying explanation of them entails that the word 'good' refers to or expresses a genuine characteristic (of actions, states of affairs, persons, and so on), and how their meaning relates to the distinctive sorts of *endorsement* that such statements commonly convey (see Analytic ethics; Emotive meaning).

The pattern exhibited in the example of value is apparent throughout philosophy. We are interested in knowledge, fiction, necessity, causation, or sensation, so we find ourselves studying statements *about* what interests us: statements attributing knowledge, describing fictions, asserting necessities, assigning causes and reporting sensations. Tools from the philosophy of language make available quite a number of views about what these statements mean and in general about how they do their expressive and communicative work; and these views inform and support philosophical positions on the real objects of philosophical interest. There have been dramatic and no doubt exaggerated claims about such techniques - for instance, that philosophy should simply *consist* in this sort of study of language. But it is if anything an understatement to say that linguistic sophistication has deepened philosophical understanding and has advanced debate in nearly all areas of philosophy (see Conceptual analysis).

See also: Adverbs; Ambiguity; Analyticity; Anaphora; Compositionality; Counterfactual conditionals; Deconstruction; Demonstratives and indexicals; Descriptions; Discourse semantics; Fiction, semantics of; Indicative conditionals; Indirect discourse; Language, Indian theories of; Logic in China; Logic, philosophy of; Logical form; Mass terms; Meaning in Islamic philosophy; Meaning, Indian theories of; Moscow-Tartu School; Proper names; Propositional attitude statements; Propositions, sentences and statements; Questions; Religious language; Scope; Semantics; Semiotics; Sense and reference; Vagueness

MARK CRIMMINS

References and further reading

Devitt, M. and Sterelny, K. (1987) *Language and Reality*, Oxford: Blackwell.(A brief and readable introduction to some central issues, with useful references and bibliography.)

Martinich, A. (1990) *The Philosophy of Language*, Oxford: Oxford University Press, 2nd edn.(A collection of many key writings - often quite difficult - in philosophy of language.)

Language, Renaissance philosophy of

Renaissance philosophy of language is in its essentials a continuation of medieval philosophy of language as it developed in the fourteenth century. However, there were three big changes in the fifteenth and sixteenth centuries. First, humanism led to a much greater interest in the practical study of languages, including Greek, Hebrew and vernacular languages, as well as classical Latin. Literary analysis and eloquent discourse were emphasized. Second, there was a loss of interest in such medieval developments as supposition theory, which meant that there was little discussion in logic texts of how words relate to each other in propositional contexts, and how sense and reference are affected by the presence of such logical terms as 'all', 'none', 'only', 'except' and so on. Only in early sixteenth-century Paris were these issues pursued with any enthusiasm. Third, the fourteenth-century insistence that both words and concepts were signs had several effects. There was a new interest in the classification of different sorts of signs, both linguistic and non-linguistic, particularly in the work of some early sixteenth-century Spaniards. Naturally significant mental language was emphasized in a way that diverted the attention of logicians from spoken languages and their imperfections. Finally, concepts themselves came in for more attention, so that many of the topics discussed by logicians overlapped with what would now count as philosophy of mind, as well as with metaphysics. For instance, philosophers in the late scholastic tradition made much use of an early fourteenth-century distinction between the formal concept, which is a representative act of mind, and the so-called objective concept, which is whatever it is that is represented by a formal concept. The discussion of these issues by such writers as Pedro da Fonseca and Francisco Suárez has an obvious bearing on developments in early modern philosophy.

1 Signs and signification

The central semantic claim of Renaissance thinkers was that the word or term is a sign, and the central semantic notion was that of signification. Frequent reference was made to Aristotle's remark (*De interpretatione* 16a3-4), 'Spoken words are signs of concepts' (see Aristotle §4); and there was widespread use of Pierre d'Ailly's definition, 'To signify is to represent some thing or some things or in some way (*aliqualiter*) to a cognitive power' (see Ailly, P. d'). Since categorematic terms such as 'cow' (unlike the categorematic terms 'concept' or 'mermaid') are intended to point to things in the world, a three-place relation of word, concept and thing was established. This account raises several problems: how syncategorematic terms (for example, 'all', 'none') can be signs; how non-referring categorematic terms are to be explained (see §3); and the precise nature of the word-concept-thing relation.

The problem of syncategorematic terms was much discussed by the nominalists of early sixteenth-century Paris (see Major, J. §2), and, following d'Ailly's definition, it was agreed that such terms signified in some way. That is, a syncategorematic term performs a function rather than pointing to an object, and this function is not itself signified by the term, though it can be in a separate locution. If I say 'No dogs are running', the act of negation is exercised; if I say 'That sentence is negative', the act of negation is signified.

The notion of signifying in some way was also extended to propositional signification by some authors, notably Fernando de Enzinas, who argued that a proposition (in the then-standard sense of an occurrent declarative sentence) does not signify things, whether ordinary things in the world or propositional complexes enjoying a special kind of existence, but performs a function. This discussion of propositional signification was linked with discussion of such issues as whether the proposition is the object of knowledge and judgment (assent and dissent) (see Toletus, F. §5), and whether it is the bearer of truth-values. These topics were still of interest to early seventeenth-century scholastics.

The precise nature of the word-concept-thing relation was the focus of the long-standing debate whether spoken words signify concepts or things. This question first became popular in the late thirteenth century, and it was still discussed by such late sixteenth- and early seventeenth-century authors as Franciscus Toletus, Sebastian de Couto (see Collegium Conimbricense) and John of St Thomas (§2). All the participants in the debate agreed on certain points. They agreed that concepts play an essential role in the significative process, for we cannot refer to objects we have no notion of. They also agreed that words are typically used to pick out things in the world rather than our own concepts. The debate concerned the way in which the role of concepts in the significative process was to be

described. Some authors, following such disparate figures as Aquinas and John Buridan, held that words primarily signified concepts and only secondarily signified things. Others, following William of Ockham, held that words signified things alone in virtue of being subordinated to concepts. Sebastian de Couto preferred to say that while the signification of things had a certain primacy, both concepts and things were made known by words, which thus enjoy a double signification. John of St Thomas held that words signify concepts more immediately and things more principally.

What has been said so far applies equally well to medieval and Renaissance treatments of the word as sign. There was, however, an important development in the fourteenth century which helps differentiate the later period from the earlier, and which stems from Ockham's insistence that the concept itself must be regarded as a sign. This notion (while foreign to Augustine) was not new, but Ockham made it central. In doing so, he made mental language (see §2) rather than spoken language the paradigm of signification.

Once the concept had come to be regarded as a sign, we find in some early sixteenth-century authors, especially Domingo de Soto (§2), a careful classification of signs, both linguistic and non-linguistic. In relation to the speaker, spoken words were said to be instrumental signs, because they were used as instruments of communication, and mental terms were said to be formal signs, because they represented by their very nature. In relation to the things signified, spoken words were said to be conventional signs, and mental terms were said to be natural signs, since their signification did not depend on choice or convention. In the various editions of his *Summulae*, Soto asked whether one should add a category of customary signs, such as napkins on the table as a sign of lunch. In the end he concluded that these signs were natural, albeit founded on a convention. Soto's classifications were elaborated by many later authors, including Sebastian de Couto and John of St Thomas; though at least one late sixteenth-century Spaniard, Domingo Báñez, did explicitly reject the view that the concept was a formal sign.

2 Mental, spoken, natural and conventional languages

The doctrine that the concept is a formal sign went hand-in-hand with the notion of mental language, a language of thought that is naturally significant and common to all human beings. Such a notion is present at least from Augustine on, but it was fully developed in the fourteenth century, first by William of Ockham and then by Pierre d'Ailly (see William of Ockham; Ailly, P. d'). Mental propositions were thought of as having syntactic structure and mental terms were thought of as having supposition, so that the notion of a language system was internalized. As a corollary, the place of grammar in the study of logic and philosophy of language was devalued. Some of the problems raised by the notion of mental language, such as the difficulty of identifying the mental correlates of demonstratives and pronouns, or of impersonal verbs such as *pluit* ('It is raining'), were still the object of lively debate in early sixteenth-century Paris. Later, the discussion of the structure of mental language disappeared, but the notion of an inner language remained. For instance, in his *Scholae (Lectures)* Petrus Ramus (§2) argued that logic or dialectic deals directly with thought, with *ratio* (reason) rather than *oratio* (discourse).

Whether mental language was seen as some kind of ideal language is unclear. Mental language certainly provided some kind of universality in that all spoken language was subordinated to it. However, it did not necessarily provide a common syntactic structure, for logically equivalent spoken propositions could turn out to be subordinated to different mental propositions. Nor was mental language necessarily thought of as being ideal in the sense of reproducing the structure of the world. On the other hand, it was supposedly ideal in the sense of being a language which was free of ambiguous terms and which contained only as much structure as was required for the formulation of judgments.

The discussion of spoken language was often related to two key episodes in the Bible: Adam's naming of the animals, and God's inflicting different languages on human beings as a punishment for building the Tower of Babel. One issue had to do with the nature of the original language and its relation to later languages. If the first language was Hebrew, as Isidore of Seville had held, in what form did it survive? If it was not Hebrew, did it completely disappear, or did it contribute at least some structure to the post-Babel tongues? A second issue had to do with the original institution of language: did Adam form it, or did God give him his language (as Sebastian de Couto, among others, held)? This in turn raised a third issue. Was the language instituted by Adam, or by God through Adam, a natural language in the sense of one that enabled users to grasp essences by virtue of a natural relationship between spoken words and the things named?

While this issue became more important in the seventeenth century with the work of such authors as Jacob Boehme and Francis Mercurius van Helmont, there was considerable discussion in the sixteenth century. The interest was due not only to biblical studies but also to the rediscovery of Plato's *Cratylus* and other classical sources, as well as to the strong Renaissance interest in magic and the Kabbalah, with the concomitant hope that a knowledge of natural language would enable one to exercise some control over the objects signified. The consensus, particularly among logicians, was that Aristotle was right, and spoken language is indeed conventional, whether it is directly God-given or not. On the other hand, the belief that language was conventional was seen as compatible with the belief that the institution of language is guided by reason (as Pseudo-Kilwardby had argued in the thirteenth century). To say that spoken words have signification *ad placitum* (by convention or agreement) does not mean that their signification is random and unmotivated. This point was emphasized by the Spanish grammarian Sanctius (Francisco Sánchez de la Brozas) in his *Minerva*, published in 1587. He felt that Aristotle, Plato and the Bible could be reconciled; and he also argued that the rational origin of language provided an underlying system of regularities in conventional languages.

3 Intentional contexts and beings of reason

Given the focus on spoken and mental terms as signs, one must ask what is referred to in intentional and modal contexts, or by fictional terms such as 'chimera'. Authors who worked within the medieval logical tradition, particularly the nominalists of early sixteenth-century Paris, attempted to treat both problems within the framework of supposition theory. There are elaborate discussions of the reference of 'horse' in such contexts as 'I promise you a horse', 'A horse is necessary for riding', 'A horse is imaginarily a chimera'. In particular, people asked whether it was legitimate to postulate reference to imaginary and impossible objects, or whether the reference of terms in special contexts could be explained solely in terms of reference to ordinary past, present, future and possible objects. In the later sixteenth century, this type of discussion disappeared from most logic texts, though there are still references in Pedro da Fonseca.

What did survive, at least in the scholastic tradition, was the discussion of beings of reason (*entia rationis*). This investigation, which can be traced back to the thirteenth century, did not focus on the problem of reference in special propositional contexts, but on the reference of terms in general. There were said to be two types of being, real beings which (with the exception of God) fall under Aristotle's ten categories, and beings of reason (*entia rationis*), which include negations, privations and relations of reason. The latter included the logical second intentions (that is, higher-order concepts used to organize first-order concepts) such as 'genus' and 'species'. Beings of reason are non-categorial and exist only because of the work of the human mind, though they may have some foundation in reality. 'Nonbeing', for instance, is a term which is significant as referring to a pure negation; and 'blindness' is a term which is significant as referring to a privation, the absence of a capacity that animals normally have; yet neither pure negations nor privations can be literally parts of the world. They must be conceptual entities. Genus and species terms (as opposed to the terms 'genus' and 'species' themselves) were often included among the terms whose significance involves relations of reason. As Domingo de Soto (§2) indicated, the point here is that when we say 'A human being is an animal' our general terms do not refer to special universal things, humanity and animality, but they imply a well-founded relation between particular individuals in the world and the concepts we use to classify them into groups. There was general agreement that chimeras should be classified as negations, and some people held that they were the most proper examples of beings of reason since, as impossible, they must be completely mind-dependent (see John of St Thomas §§2, 4; Suárez, F. §2).

4 Analogy

Analogy as a theory about a certain sort of linguistic usage has thirteenth-century roots, and involves three problems. First, there is the problem of equivocal terms as Aristotle introduces it at the beginning of his *Categories*. Equivocal terms (for example, 'bank') are those which can be used in two quite different senses, and it seems natural to extend the notion of an equivocal term to cover terms that are used in different but related senses. Second, there is the metaphysical problem of how to discuss being (*ens*) in general when the being of a substance is so different from the being of an accident. Third, there is the problem of religious language. How can words normally used of humans, such as 'just' or 'good', be meaningfully used of God, when God is so different from human beings? In answer to these problems, logicians and theologians developed a theory which divided words into three sorts, independently of context. Some were univocal (always used with the same sense), some were

purely equivocal (used with totally different senses), and some were analogical (used with related senses).

The term 'analogy' itself had two senses. In the original (Greek) sense, it involved a comparison of two proportions or relations. Thus 'principle' was said to be an analogical term when said of a point and of a spring of water because a point is related to a line as a spring is related to a river. This type of analogy came to be called the analogy of proportion, proportionality, or (by Cajetan) proper proportionality. In the second sense, analogy involved a relation between two things (or one pair of things and a third), of which one is secondary and the other primary. Thus 'healthy' was said to be an analogical term when said of a dog and its food because while the dog has health directly, its food is healthy only as contributing to or causing the health of the dog. This second type of analogy became known as the analogy of attribution.

One of the main subjects of debate was how to classify types of analogy, and how to apply the various types to the different metaphysical and theological cases mentioned above. Although in *De veritate (On Truth)* Aquinas said that religious language must be interpreted by means of the analogy of proportionality, in other writings he appealed to the analogy of attribution. During the fourteenth and fifteenth centuries most logicians and theologians, including Capreolus, appealed to the analogy of attribution (if they discussed the topic at all). At the end of the fifteenth century, however, Thomas de Vio, Cardinal Cajetan argued that the analogy of proportionality was the only true analogy (see Cajetan §2). Cajetan's view was not always accepted, but there was much discussion of the issue by the later scholastics, including Soto, Fonseca (§3) and Suárez (§2) (see Silvestri, F.).

The other main subject of debate is very closely related to philosophy of mind, and springs mainly from the work of early fourteenth-century philosophers, particularly Duns Scotus. Latin translations of Aristotle maintained that the difference between a univocal term and an equivocal term was that the latter was subordinated to more than one *ratio*. The *ratio* soon came to be identified with a concept, and the question then was how many concepts are involved when an analogical term is used. There were three views among logicians and theologians. Analogical terms could be seen as straightforwardly equivocal terms subordinated to two distinct concepts; they could be seen as subordinated to an ordered cluster of concepts (possibly but not necessarily described as a disjunction of concepts), or they could be subordinated to a single concept which represents in both a prior and a posterior manner (*per prius et posterius*). Scotus and his followers rejected all these possibilities, and argued that the word *ens* was univocal, but a lively discussion continued (especially among Thomists) until the time of Suárez.

Many of these thinkers made a distinction between formal and objective concepts. The formal concept was the act of mind or conception that represented an object, and the objective concept was the object represented. If the spoken word 'being' corresponds to just one formal concept (a point on which there were some differences of opinion), the focus of discussion shifts to the status of the objective concept. Is it the actual thing in the world which is thought about; is it a common nature or some other kind of intermediary entity which is distinct from the external object without being mind-dependent; or is it a special kind of mind-dependent object which has only objective being, the being of 'being thought' (*esse objective*, *esse cognitum*)? It is at this point that philosophy of mind and ontology take over from philosophy of language.

5 Humanism

Humanists, including Valla, Erasmus and Vives, were more concerned with the description of language as a literary phenomenon and with the cultivation of eloquence than with philosophy of language proper. The late fifteenth and sixteenth centuries saw an explosion of practical language studies. New humanist grammars appeared which prepared students for literary analysis rather than philosophy and theology; Greek and Hebrew studies expanded; there was an increasing use of and interest in vernacular languages; biblical studies and translations flourished; the widespread use of printing focused attention on problems of spelling and grammatical norms; and there was new interest in forms of writing, including Hebrew characters and Egyptian hieroglyphs. This last interest was often combined with a belief in magic. Agrippa, for instance, discussed how one might manipulate the numerical values of the letters of the Hebrew alphabet in order to make contact with the spirit world.

See also: Humanism, Renaissance; Language, medieval theories of; Logic, medieval; Logic, Renaissance; Universal language

E.J. ASHWORTH

References and further reading

Ashworth, E.J. (1985) *Studies in Post-Medieval Semantics*, London: Variorum.(A collection of articles including studies of reference in intentional contexts, propositions and mental language.)

Ashworth, E.J. (1988) 'Traditional Logic', in C.B. Schmitt, Q. Skinner and E. Kessler (eds) *The Cambridge History of Renaissance Philosophy*, Cambridge: Cambridge University Press, esp. 153-62.(A useful general account with plenty of references.)

Ashworth, E.J. (1990) 'Domingo de Soto (1494-1560) and the Doctrine of Signs', in G.L. Bursill-Hall, S. Ebbesen and K. Koerner (eds) *De Ortu Grammaticae: Studies in Medieval Grammar and Linguistic Theory in Memory of Jan Pinborg*, Amsterdam and Philadelphia, PA: John Benjamins, 35-48.(Discussion of how signs were classified.)

Ashworth, E.J. (1995) 'Suárez on the Analogy of Being: Some Historical Background', *Vivarium* 33: 50-75.(How analogical terms were classified; formal and objective concepts; bibliography.)

Coudert, A. (1978) 'Some Theories of a Natural Language from the Renaissance to the Seventeenth Century', in *Studia Leibnitiana, Sonderheft 7*, Wiesbaden: Franz Steiner, 56-114.(A clear overview.)

Demonet, M.-L. (1992) *Les Voix du signe: Nature et origine du langage à la Renaissance (1480-1580) (Voices of the Sign: the Nature and Origins of Language in the Renaissance)*, Paris and Geneva: Champion-Slatkine.(A full, fascinating account of all aspects of language based on both philosophical and literary sources.)

Jardine, L. (1988) 'Humanistic Logic', in C.B. Schmitt, Q. Skinner and E. Kessler (eds) *The Cambridge History of Renaissance Philosophy*, Cambridge: Cambridge University Press, 173-98.(A useful general account, though controversial on Lorenzo Valla; shows how humanist logic was used in textual exegesis.)

Nuchelmans, G. (1980) *Late-Scholastic and Humanist Theories of the Proposition*, Amsterdam, Oxford and New York: North Holland.(Discusses mental language, the significate of the proposition, and the terminological innovations of Melanchthon and Ramus.)

Nuchelmans, G. (1983) *Judgment and Proposition from Descartes to Kant*, Amsterdam, Oxford and New York: North Holland.(First chapter discusses Renaissance Thomism, including material on formal and objective concepts.)

Padley, G.A. (1976) *Grammatical Theory in Western Europe, 1500-1700. The Latin Tradition*, Cambridge: Cambridge University Press.(Generally good, but to be used with caution where logic is concerned.)

Percival, W.K. (1982) 'Changes in the approach to language', in N. Kretzmann, A. Kenny and J. Pinborg (eds) *The Cambridge History of Later Medieval Philosophy*, Cambridge: Cambridge University Press, 808-17. (Discussion of humanist grammar teaching.)

Roncaglia, G. (1995) 'Smiglecius on *entia rationis*', *Vivarium* 33: 27-49.(Useful introduction to the notion of beings of reason, with good bibliography.)

Language, social nature of

Language is mostly used in a social setting. We use it to communicate with others. We depend on others when learning language, and we constantly borrow one another's uses of expression. Language helps us perform various social functions, and many of its uses have become institutionalized. But none of these reflections settle the question of whether language is an essentially social phenomenon. To address this we must consider the nature of language itself, and then ask which social elements, if any, make an essential contribution to its nature.

While many would accept that language is an activity that must take place in a social setting, others have gone further by arguing that language is a social practice. This view commits one to the claim that the meanings of an individual's words are the meanings they have in the common language. The former view need not accept so strong a claim: meaning depends on social interaction because it is a matter of what one can communicate to others but this does not require the existence of communal languages. A competing conception which rejects the social character of language in either of these versions is the thesis that language is mentally represented in the mind of an individual.

1 Linguistic Platonism

Languages can be used in many ways and for many purposes, but a dispute about the nature of language concerns language in its most fundamental sense. Is it fundamentally social or do individual speakers have their own languages?

One view of language, which will not be discussed in any detail in this entry, is linguistic Platonism. This is the view that languages are formally characterized abstract entities existing independently of all speakers. Platonists then seek to define a relation between these abstract objects and speakers to determine which is the actual language of an individual or a population. The actual language relation may be defined either in terms of the conventional practices adopted by a population, or the psychological make-up or linguistic intuitions of the individual language-user. Thus Platonism does not settle the issue of whether languages are properties of individuals or social groups.

2 Communal languages and idiolects

A broad division exists between those philosophers and linguists who think the fundamental conception of language is that of a shared public language, and those who consider the fundamental notion to be that of an individual's language or idiolect. Within each of these schools there is room for disagreement about the precise nature of languages and idiolects. These differences will be reviewed below.

Michael Dummett (1989), following the ideas of Ludwig Wittgenstein (1953), offers a number of arguments for the communal conception of language. One of these goes via his rejection of Platonism. For Dummett, the objectivity of meaning requires the meaning of a word to be independent of any given speaker's judgment of it. But if the meaning of words were independent of all speakers, as the Platonist suggests, this would open up the sceptical possibility that all speakers could be in ignorance about the actual meaning of words in their language. Rejecting this possibility, Dummett insists that what an expression means among a community of language-users must be determined by the publicly observable conventions governing its use.

Dummett also argues that the publicity of meaning is required to ensure successful communication. Because speakers hold themselves responsible to the conventional meanings of words in taking themselves to be speaking a given language, what they say when uttering certain words is what those words mean in the communal language to which they belong. It is only because there are such conventional practices that meaning is shared and communicable; and for Dummett, were it not for such standards of use, speakers could not know they attached the same significance to their words. Meanings would be private and incommunicable, hidden in the minds of each individual speaker. But if this were the case, we could never learn a language, or know if someone else understood it. But we can work out what people mean by the observable use they make of their expressions; and to learn or understand a language is simply to participate in these shared linguistic practices. However, no speaker will be fully competent in the conventions governing the use of words and rules for grammatically compounding them; at most speakers will have a partial mastery of the public language. Nevertheless, for Dummett communication

depends on the standards of the communal practice.

Similar considerations have been advanced by John McDowell (1988) who regards the possibility of thought and of knowing the minds of others as depending on our speaking a communal language. We arrive at our thoughts by finding the words to express them, and so in virtue of the language in which we do this we make our minds immediately available to one another. The words you use to express a thought will express the same thought for me and so, just in virtue of hearing you speak sincerely, I can know what you think. This is possible because the meanings of a speaker's words lie open to view on the surface of his practice. But they are recognizable only to those who share those practices. These ideas hail from the later work of Wittgenstein, who saw language as a social practice governed by rules of use (see Meaning and rule-following). Along with Wittgenstein, Dummett and McDowell assume that in the absence of shared languages, or observable conventions of use, our knowledge of others' meanings would be a matter of psychological speculation about what goes on in a private realm concealed behind behaviour. And, given the doubtful coherence of a private language, McDowell thinks the absence of shared languages would de-stabilize communication, meaning and thought.

Not every defender of communal languages accepts the publicity of meaning. It is possible to argue that private states are involved in the meanings of public language items so long as speakers have sufficiently similar responses to the situations in which those words are used. On this view, it would be enough for the purposes of communication that speakers attached the same significance to terms in the language: they would not also have to know that they did. Success in communication would be a contingent matter and could never be guaranteed (see Private states and language).

The strongest opposition to the communal view of language comes from those advocating the idiolectic conception of language, although there is deep disagreement amongst the protagonists over the extent to which an idiolect is social or psychological in nature. However, all proponents of the idiolectic view resist the assumption made by Dummett and McDowell that if a language is not shared then the meaning of an individual's words would have to be subjective and private. It does not follow from the fact that meanings are publicly discoverable that they are shared. I can work out the meaning someone attaches to a word without attaching that meaning to it myself. Donald Davidson (1986) stresses malapropisms as providing cases of this kind (see §4). The denial of shared language does not entail the privacy of meaning and the publicity of meaning does not require a communal language. The arguments of McDowell and Dummett designed to show that the language of an individual would be necessarily private remain inconclusive. Facts about my idiolect may be knowable by those I interact with, and linguists studying me scientifically may know more about my idiolect than I do.

3 Language as a psychological or a social phenomenon

Noam Chomsky (1986) is a vigorous opponent of the notion of shared public languages. He argues that there is no empirically respectable way to define such a notion. According to Chomsky there is no such thing, for example, as English, or Dutch, or Chinese. In each case there is no single set of grammatical rules, shared vocabulary items, or rules for pronunciation which characterize the language in question. Whether we include every idiosyncratic use of every word and rule of grammar by a particular population of speakers, or select only the intersection of those uses, what is specified is a language nobody knows or speaks. The same problems beset attempts to delimit a dialect. For this reason, Chomsky abandons the everyday use of 'language', which he calls E-language: this being an ill-assorted set of overlapping practices, patterns of deference, power relations among speakers, and so on. He replaces it with the notion of a speaker's I-language: an internal component of the individual's mind/brain which assigns meaning and structure to the sounds and signs the speaker encounters. The I-language is a cognitive psychological entity - a state of the language faculty - which together with a lexicon of word-like items determines the extent of the speaker's idiolect.

According to Chomsky, language is essentially defined by its meanings and structures, but these exist only as mental representations in the mind of the speaker - as part of the speaker's knowledge of language - and have no reality in the external or social world. Linguistically, the world contains only sounds and marks: it is creatures equipped with a language faculty who assign them meaning and structure. At first, this is achieved without reference to others. Our earliest interactions will of course influence our acquisition of vocabulary and the particular grammar we attain although, in the case of the latter, Chomsky regards this as environmental triggering: a prompt from one's surroundings to set the parameters of a system that is already in place. Thus the development

of one rather than another I-language is the result of interaction between the linguistic environment and the innate linguistic system (see Language, innateness of).

After we attain mature linguistic competence, we often defer to others in positions of authority in their use of words and assessments of grammar. But Chomsky points out that this is social behaviour undertaken after we have acquired a language, and so it does not constitute an essentially social contribution to it. On the notion of language Chomsky favours, the study of language becomes a branch of individual cognitive psychology. The study of group behaviour has no impact on it. Communication is not, on this picture, essential to language. Language is primarily a means of expressing thought. Speakers and hearers do have the means to try to make sense of one another's talk and the closer their I-languages and vocabularies are to one another, the more chance they will have of success. However, for Chomsky, communication is risky and success is never guaranteed so it cannot provide the basis for linguistic competence.

Where he is willing to acknowledge a pervasive social aspect of language is in the case of the division of linguistic labour: a phenomenon Hilary Putnam (1975) drew attention to. This occurs where there are words having an everyday sense and a more technical meaning without being ambiguous. The words 'gold' and 'carburettor' provide examples for those with little knowledge of chemistry or cars. Speakers can use these words without knowing their precise meanings because they defer to experts in their community who know how to fix the meaning and reference of those terms precisely (see Putnam, H. §3). Chomsky points out that the entries for such words in a speaker's mental lexicon will specify the meaning of the terms as far as the speaker's (partial) knowledge carries him, but they would also include an indication that further details are to be filled out by others. The meanings specified in this way do not go beyond what can be studied in an individual's knowledge of language. It is not to concede that language includes any essentially social elements. The phenomenon Putnam alludes to, while extensive, is not fundamental to the use of language. People could cease to use terms they did not fully understand without ceasing to speak or understand a language. What this shows is that the division of linguistic labour provides the weakest claim for the social character of language. Other reasons are needed to conclude that language is necessarily social in character.

4 Idiolects as social phenomena

Arguments for this are provided by Donald Davidson and Tyler Burge, although they differ in their conceptions of what makes an idiolect social.

Davidson rejects the idea of language as 'a clearly defined shared structure' which language-users learn and deploy in communicative exchanges. The appeal to conventions is neither necessary nor sufficient for linguistic communication, since speakers and hearers need not attach the same meanings to their words nor stick to the conventions they have established in order to make sense of one another's remarks. All that is required is that the hearer know what the speaker means; it is not additionally required that the hearer attach the same significance to their own use of those words. The cases of malapropism and slips of the tongue serve to make the point. I can know that someone uses 'debated' as I use 'abated' when they say 'The rain has debated'. But there is no reason for me to adopt their use of that word. So long as I can work out, by means of interpretation, what significance the speaker intends their words to have, I can understand them without having to share their language. Communication does not require speaker and hearer to belong to the same linguistic community, or to share a set of linguistic practices; it requires them to share an ability to interpret one another's utterances as part of a larger project of making sense of one another's behaviour in rational terms through the attribution of beliefs, desires and intentions. Part of this project will consist in assigning meanings to people's utterances that make sense of what they say against the background of the mental states we take them to be in. Clues of all kinds can figure in the evidence an interpreter draws upon to work out the meaning of someone's remarks, and no single set of shared conventions will be necessary or sufficient for this task.

While it is unlikely that any two speakers will ever speak the same language - minor differences in their vocabularies or grammar will always distinguish their idiolects - what they must share, according to Davidson, is the method of interpreting one another's speech. It is this condition which ensures the notion of idiolect has a social dimension since speakers can only mean what others can recognize or interpret them as meaning. The meanings of the words and sentences of their idiolects are fixed by others' publicly determined interpretations. For Davidson, the possibility of meaning anything depends on successful communication with another: having one's

words interpreted as one intends them to be interpreted. In this way, meaningful speech, although it demands nothing more than a variety of idiolects, is necessarily social (see Radical translation and radical interpretation §§7-10).

Burge (1989), in contrast, argues that social factors beyond the scope of local communicative interactions play an essential role in the meaning of words in the individual's idiolect. He takes language to be a partly social phenomenon not just because it is learned and used in a social setting but because the meanings of a person's words depend on the way others in their linguistic community use them. He agrees with Chomsky that the study of language is a project for individual psychology, but he argues the semantic facts about an individual's idiolect depend essentially on their relations to other language-users. To establish that meaning must be individuated nonindividualistically and that idiolects are to some extent social, Burge (1982) constructed a famous thought-experiment in which there is an individual who uses the word 'arthritis' to express a number of thoughts about pain in their joints. One day they declare 'I have arthritis in my thigh'. Since the community the individual belongs to uses the word to apply only to inflammation of the joints, what the individual says is false. Burge then asks us to consider a counterfactual situation in which the community the subject belongs to uses 'arthritis' to apply both to inflammations of the joints and to other rheumatoid ailments. In this counterfactual situation there is no change in the individual's physical history or nonintentionally characterized experiences, and yet the statement they make there is true not false. Burge invites us to share the intuition that the truth-value of what is said differs in the two situations because something different is meant in each context. He concludes that meaning of the word 'arthritis' on the speaker's lips is different in each of these situations because of the different linguistic communities to which they belong. By this argument Burge hopes to show that the meanings of someone's words are individuated not just by facts about the individual but also partly by facts about the uses of words in the wider linguistic community (see Content: wide and narrow). In this way, the meanings of many words in a person's language are not just up to him but depend essentially on the linguistic practices of others around him. But what is it for a speaker to belong to a particular linguistic community? For Burge it is a matter of his deferring to members of that community to explicate the meanings and determine the reference of his words on what are cognitive and not just pragmatic grounds. If Burge is right, facts about the meaning of our words supervene not only on facts about our use of words but also on facts about other people's usage. This amounts to a strong claim that social factors are constitutive of meaning and hence of language and that, although language is a matter of individual psychology, it is also partly a social phenomenon.

See also: Language, conventionality of; Methodological individualism; Language and gender; Meaning and communication

BARRY C. SMITH

References and further reading

Burge, T. (1982) 'Individualism and the Mental', *Midwest Studies* 4: 73-121.(The source of the famous thought experiment discussed in §4, containing objections to it and responses.)

Burge, T. (1989) 'Wherein is Language Social', in A. George (ed.) *Reflections on Chomsky*, Oxford: Blackwell, 1989.(A demanding paper which contains his exposition of the partly social, partly psychological conception Burge has of idiolects.)

Chomsky, N. (1986) *Knowledge of Language*, New York: Praeger.(Chapters 1 and 2 contain a clear statement of Chomsky's conception of I-language discussed in §3.)

Davidson, D. (1986) 'A Nice Derangement of Epitaphs', in E. Lepore (ed.) *Truth and Interpretation*, Oxford: Blackwell, 1986.(An important paper challenging the need for conventions and rules of language, but defending the social nature of communicative exchanges.)

Dummett, M.A.E. (1989) 'Language and Communication', in A.George (ed.) *Reflections on Chomsky*, Oxford: Blackwell, 1989.(Careful discussion of many of the issues raised in §§2-3, including critical discussions of Davidson and Chomsky.)

McDowell, J. (1988) 'In Defence of Modesty', in B. Taylor (ed.) *Michael Dummett: Contributions to Philosophy*, Dordrecht: Martinus Nijhoff, 1988.(A subtle and penetrating account of McDowell's views on language and mind.)

Putnam, H. (1975) 'The Meaning of "Meaning"', in *Collected Papers*, vol. 2, Cambridge: Cambridge University Press.(A seminal paper containing the notion of the division of linguistic labour among many other influential ideas.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(A difficult read but taken to be the origin, and the most powerful sustained defence of, the philosophical conception of language as a social practice.)

Linguistic discrimination

'Linguistic discrimination' is a redundancy. Discriminating is at the heart of what languages do. The question, of course, is when they can be said to do it invidiously, or rather when we, in our use of language, can be said to be discriminating invidiously. In Aristotelian terms, the proper use of linguistic discriminations is to make the right sort and number of discriminations in the right ways and at the right times - that is, not to discriminate between those things that, for the legitimate purposes at hand, ought to be seen as the same; to discriminate between those things that, for the legitimate purposes at hand, ought to be seen as different; and to discriminate in ways that advance legitimate and not illegitimate purposes. Disputes about what constitutes linguistic discrimination (in the invidious sense) revolve around both the legitimacy of our purposes and, in the light of those purposes, the aptness of particular discriminations. Such disputes presume both that our language shapes our actions (that linguistic discrimination plays a role in maintaining unjust inequalities) and that our actions can shape our language (that acknowledging such discrimination can and should lead to linguistic change).

1 Arguments for causality and remediability

Claims for the causal efficacy of linguistic discrimination range from the claim that some thoughts are made either unavoidable or impossible because of the language in which thinking occurs, to the claim that language shapes, without determining, how we think. The stronger claim, often appealing to the Sapir-Whorf hypothesis, sees language as providing the conceptual tools without which the experienced world is inchoate and only in terms of which can sense be made (see Sapir-Whorf hypothesis). In its strongest form, this hypothesis is surely false: language-users are inventive, and all languages have room for invention.

It is harder to refute the view that the languages we use shape our thoughts, making some ways of thinking more 'natural', more readily comprehended by our interlocutors, harder to avoid if they come to seem problematic. Richly nuanced vocabularies exist to describe some parts of the experienced world and not others. Nor are all the differences semantic: Black English is less well suited than is Standard English for expressing the agentless passive, that is, the idea that 'it happened' without indicating that someone or other did something (Jordan 1985).

That the language we use does not absolutely constrain us is no reason not to attend to the ways in which it shapes our thinking, leading us to imply what we do not intend, or making it difficult for us to say what we mean, or what we would mean if we had the words to mean it with. It is also remediable. Languages constantly change in response to an indefinite range of social forces; and while there can be no guarantee that a suggested reform will catch on, there is no reason a priori to argue that it could not do so. One may choose to resist it, but that choice is as political as is the argument for change.

2 Taxonomic schemes

Metaphysicians have long debated the nature of kinds. Are some of them 'natural' (whatever that means)? Are only such kinds, if there are any, 'real'? What about kinds created by our taxonomizing practices? Inspired largely by the work of Michel Foucault, theorists are examining the social and political contexts within which taxonomies operate and the value judgments they both depend on and reinforce (see Root 1993). Case studies explore the social construction of race (Goldberg 1990) and gender (Lorber 1994); as well as of kinds of practices such as sexual harassment (MacKinnon 1979) and child abuse (Hacking 1992), conditions such as premenstrual syndrome (Zita 1988), life stages such as childhood (Aries 1962), and persons such as homosexuals (Stein 1992) and heterosexuals (Katz 1990). Questions about invidious discrimination arise with respect to kinds that arguably rest on dubious science or on politically questionable social practices.

Thus racial classifications have been argued to be discriminatory because they are biologically unfounded, on the assumption that races, if 'real', are so because of discoverable biological differences between different racial populations. Since there appear not to be such differences, racial classifications distinguish as different those who, on the alleged basis, are in fact the same. Alternatively, it can be argued that race is best understood as a social classification, in which case the issue is whether as such race can be made coherent and, if so, whether the social practices that make it coherent are just. If racial classification cannot be made coherent, then it is discriminatory for distinguishing between people who, on its own terms, are not distinguishable; if it is coherent but grounded on

unjust social practices, then racial classification would be invidiously discriminatory because it was wrong to engage in the practices that make it descriptively correct. Others argue that, although founded on racist practices, racial classification has acquired a conceptually adequate grounding in practices that include those of anti-racist resistance and positive group identification (see Race, theories of). Similar questions are raised about the classifying of persons as homosexual or heterosexual: apart from social practices of stigmatizing homosexuality as unnatural, perverse or sick, what reasons are there for classifying persons according to whether they desire those of the same or the opposite gender (see Sexuality, philosophy of)?

The situation is somewhat different regarding discrimination by gender (see Language and gender). There is better reason than in the case of race to regard the distinction between males and females as grounded in biology, although it is a matter of dispute whether the social practices that at the very least build upon, extend and 'clean up' the biological differences are morally and politically justifiable. But there is a further question: gender ascription, in English and in many other languages, is uniquely non-optional, in particular because of the gender inflection of pronouns. This ubiquitous marking of gender is arguably discriminatory, implying that, unlike all the other equally real differences between people, gender is ubiquitously relevant. (Similarly, the argument for the use of 'Ms' rests not on the wrongness of distinguishing between unmarried and married women but on the wrongness of requiring that such a distinction be marked whenever a woman is referred to by title.) Ubiquitous gender-marking is part of the social construction of gender, both in making gender neutrality difficult to express even when it is agreed to be appropriate and in making gender ambiguity socially stigmatizing even when it is acknowledged to be possible.

3 The paradigmatic and the generic

Semantic spaces are complex: not all unquestionable members of a class are equally paradigmatic. It seems to be a feature of human cognition to learn categories by grouping members around paradigmatic exemplars, and variation in paradigms is not for the most part idiosyncratic. It is a matter of dispute how much of the contouring of semantic space is 'hard-wired', but it is clear that for at least many categories, paradigmatic exemplars owe that status to culturally variable, frequently unjust, social arrangements.

Thus, it has been pointed out (Spelman 1988) that within racist cultures, white people are paradigmatic and people of colour are variously 'different', so that even anti-racist arguments claim that for all relevant purposes, people of colour are 'just like' white people. Similarly for other forms of privilege: those who are, for example, male or middle-class or heterosexual are paradigmatic in the sense that their achievements, interests and needs are taken as the measures against which those of others are evaluated. When, however, the category is itself a subordinate one, paradigmatic exemplars are likely to be members of a relevantly subordinated group. In either case, those who conform to the paradigm are typically 'unmarked', while those who are not paradigmatic are 'marked': 'male nurse', 'gay author'.

It is a matter of controversy the extent to which such language is itself discriminatory, rather than just accurately reflecting reality. The arguments for claiming specifically linguistic discrimination point out that being in the majority and being paradigmatic are not always the same thing and, even when they are, linguistic discrimination helps to maintain, rather than innocently reflecting, social discrimination.

A limiting case of the paradigmatic is the false generic. False generics are terms for the paradigmatic members of a group improperly used to refer to all members. It has been a matter of controversy whether in English masculine nouns and pronouns are also, ambiguously, genuinely generic, or whether they should be regarded as false generics. The arguments for the latter position are both linguistic and political. Linguistic arguments (see Mercier 1995) point to the often irresolvable ambiguities that result from the employment of the masculine as generic, as well as to the anomalous and illogical rules that govern such employment. Political arguments point to the advantages that accrue to those who are taken to be generically human, rather than 'different', and to the alienation increasingly experienced by female listeners and readers uncertain of the extent to which, if at all, they are included among the 'men' under discussion.

4 Stereotyping and stigmatizing

Another type of linguistic discrimination concerns the vocabulary used to refer to members of privileged and

subordinated groups and, by analogy, parts of the nonhuman world that are variously associated with them. Tropes of light and darkness, for example, are central to European discussions of reason, especially those that were articulated concurrently with the colonization of Africa - the 'dark continent' - to which Europeans had the right and duty to bring 'enlightenment'. Gender stereotypes permeate talk about 'mother' nature, the 'hard' versus the 'soft' sciences, 'active' reason versus the 'passions'. Such vocabulary relies on and reinforces stereotypes that are inextricably bound up with socially discriminatory practices and, for reasons similar to those given in §3, can be argued to constitute linguistic discrimination.

Stigmatizing vocabulary, whether used to refer directly to members of stigmatized groups or derivatively to things associated with such groups, exemplifies linguistic discrimination not so much in how the group in question is being distinguished but in how language is being used in reference to them. The clearest examples are derogatory terms applied on the basis of race, gender, sexuality, disability, national origin, religion, and so forth. A related form of linguistic discrimination is the use of terms referring to stigmatized groups as negative characterizations: 'Indian-giver' or 'Jewing someone down'. Implicitly or explicitly gendered vocabulary typically encodes problematically stereotypical views concerning, for example, the different valuations of male versus female sexuality and assertiveness. Language referring to disabilities is often problematically generalized in ways that imply, for example, that blindness stands in the way of 'seeing' what follows from a set of premises.

Even when such forms of linguistic discrimination are not blatantly bigoted, there is good reason to avoid them, given the usual desirability of communicating as clearly as possible what is intended to the intended audience. Increasingly, audiences are reading discriminatory intent in the use of discriminatory language (including the supposedly generic masculine), and such reading habits are not in the power of authors to circumvent. To the extent that it is no part of authorial intent either to communicate negative or marginalizing attitudes towards members of particular groups, or to exclude members of such groups (and their allies) from one's audience, it is wise to eschew vocabulary that will predictably have those effects.

5 Value-laden terminology

Whether or not it is in theory possible for the discriminations made by some particular language (or part of a language) to reflect only morally and politically non-tendentious judgments of similarity and difference, it is clear that such neutrality does not characterize a large number of the distinctions made in the languages we actually use. Public discussions are significantly shaped, in ways that frequently elude explicit debate, by the language in which the issues are framed. A central tenet of moral and political thinking is that like cases ought to be treated alike; consequently the crux of many arguments lies in judgments of similarity and difference, carried out in language that can make some of those judgments nearly automatic and others virtually unthinkable.

Examples abound, from military euphemisms that draw attention towards the achievement of strategic objectives and away from killing people, to abortion debates that pit those who favour choice against those who would coerce child-bearing or those who are pro-life against those who would kill babies. Mainstream press coverage in the USA distinguishes between freedom-fighters and terrorists in ways that map not the activities in which those individuals and groups engage so much as their congruence with or opposition to US foreign policy. Large-scale violence in inner city communities is thought of in one way when called a 'riot', in another when called an 'uprising'.

The notion of an 'essentially contested concept' (Gallie 1956) raises similar questions. Such concepts are deeply value-laden, carrying strongly positive or negative connotations: 'democracy' or 'repression'. Disputants typically share a small core of paradigmatic exemplars and differ over which additional examples are relevantly similar. A discriminatory use of an essentially contested concept would be one that applied and withheld the term in ways that are argued to be either undermotivated or motivated in problematic ways.

The point is not that we ought to eschew language that encodes controversial judgments of similarity and difference. Rather, we should be aware of what is encoded in our language and ready to argue for the judgments expressed therein and to challenge those of others with which we disagree. To call a particular use of language discriminatory goes beyond the claim that it implies a disputable judgment: rather, it commits one to arguing for why and how that judgment ought to be disputed.

6 Languages and dialects

The historical development of languages has gone on in tandem with the histories of the formation of nations, the spread of and resistance to colonization and imperialism, the globalization of culture and the influence of nationalist movements, and the development and spread of communication technologies. Language is typically bound up with national identity, often in complex and contradictory ways. For example, movements of African nationalism sometimes embrace the language of the colonizer as the language of national unity, rather than divisively choosing among a number of indigenous languages.

Questions about whether something is a language, rather than a dialect, or a dialect, rather than a degraded use of language, will always raise contentious issues concerning present usage and users, as well as about their histories. Charges of discrimination will be appropriate when such judgments appear to be ethnocentric or otherwise biased, as, for example, in the controversy over whether American Black English is a legitimate dialect. The argument that it is rests on pointing out the rule-governed nature of its syntactic, semantic and phonological differences from Standard English, and on the history of its development from the interaction of West African languages with English (Smitherman 1977).

Languages and dialects are attached to particular groups of people, who are related to other groups in frequently hierarchical ways, and those hierarchies are reflected in the ways in which the languages and dialects are thought of and treated. Attempts are sometimes made to maintain the 'purity' of a language by codifying a version of it that carries both historical and present day prestige. Educational systems typically single out certain dialects as 'standard' or 'received': as the paradigmatic exemplars of the language they tend to be regarded not as dialects, but as the language itself, unmarked. Similarly, some speech patterns count as 'accents', and different accents lead to different stereotypical ascriptions of intelligence and social standing.

Reflecting the dominance of the USA following the Second World War, English is increasingly the global language - used, for example, in commerce, electronic communications, aviation and science - a development that privileges those who speak it fluently and marginalizes or threatens the survival of other languages. To regard the globalization of English as discriminatory is to regard that situation as other than natural, inevitable or equitably beneficial. Discrimination is also an issue when it comes to the semantic, syntactic and phonological diversity of English as spoken in different countries. Britain and the USA each have large amounts of rather different sorts of clout when it comes to determining whether or not something counts as 'proper' English, while even native English speakers from elsewhere are generally regarded as speaking either incorrectly or in a substandard dialect.

See also: Discrimination

NAOMI SCHEMAN

References and further reading

Aries, P. (1962) *Centuries of Childhood*, London: Jonathan Cape.(On the 'invention' of childhood. One of the earliest discussions of how an apparently natural kind - children - can be socially constructed.)

Christensen, F.M. (1994) '"Sexual Harassment" Must Be Eliminated', *Public Affairs Quarterly* 8: 1-17.(Argues against the legitimacy of the concept of sexual harassment.)

Gallie, W.B. (1956) 'Essentially Contested Concepts', *Proceedings of the Aristotelian Society* 56: 167-98. (Explanation of the concept referred to in §5.)

Goldberg, D.T. (ed.) (1990) *Anatomy of Racism*, Minneapolis, MN: University of Minnesota Press.(Includes essays on the nature of racial classification and a bibliography.)

Hacking, I. (1992) 'World-making by Kind-making: Child abuse for example', in M. Douglas and D. Hull (eds) *How Classification Works: Nelson Goodman among the Social Sciences*, Edinburgh: Edinburgh University Press.(This volume includes other relevant essays on the nature of kinds in relation to Goodman's work on similarity.)

Jordan, J. (1985) 'Nobody Mean More to Me than You and the Future Life of Willie Jordan', in *On Call: Political Essays*, Boston, MA: South End Press.(Discussion of the expressive resources of Black English and of the practical choices about its use.)

Katz, J.N. (1990) 'The Invention of Heterosexuality', *Socialist Review* 20: 7-34.(Argues that heterosexuality as an identity or 'sexual orientation' - no less than homosexuality - is distinctively modern.)

Lorber, J. (1994) *Paradoxes of Gender*, New Haven, CT: Yale University Press.(Critical overview of the social

construction of gender; includes extensive bibliography.)

MacKinnon, C.A. (1979) *Sexual Harassment of Working Women*, New Haven, CT: Yale University Press. (Germinal text for the recognition of sexual harassment as a distinctive type of activity.)

Matsuda, M.J., Lawrence III, C.R., Delgado, R. and Crenshaw, K.W. (1993) *Words that Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*, Boulder, CO: Westview Press.(Arguments for the assaultive power of language and about the legal consequences of recognizing that power, especially in light of the US constitutional protection of the freedom of speech.)

Mercier, A. (1995) 'A Perverse Case of the Contingent A Priori: On the logic of emasculating language (A reply to Dawkins and Dummett)', *Philosophical Topics* 23: 221-59.(Arguments, some linguistically technical, against the allegedly generic use of masculine language.)

Root, M. (1993) *Philosophy of Social Science*, Oxford: Blackwell.(Chapter 10, 'Sorting data into kinds', discusses the ways in which kinds can be socially constructed.)

Shapiro, M. (ed.) (1984) *Language and Politics*, New York: New York University Press.(Includes Foucault's 'The Order of Discourse' as well as a number of other, methodologically diverse essays on the language of politics and the politics of language.)

Smitherman, G. (1977) *Talkin and Testifyin: The Language of Black America*, Detroit, MI: Wayne State University Press.(Arguments for the legitimacy of Black English, an account of its history, and outlines of its syntax, semantics and pragmatics.)

Spelman, E.V. (1988) *Inessential Woman: Problems of Exclusion in Feminist Thought*, Boston, MA: Beacon Press.(Includes a discussion of the ways in which dominant (in this case white) members of a group (women) come to be the paradigmatic or even the sole understood referent of the generic term.)

Stein, E. (ed.) (1992) *Forms of Desire: Sexual Orientation and the Social Constructionist Controversy*, New York and London: Routledge.(Philosophically framed collection of essays on what it means, and whether or not it is true to say that, in particular, homosexuality as an identity or 'sexual orientation' is distinctively modern. Includes a bibliography.)

Young, A. (1995) *The Harmony of Illusions: Inventing Post-traumatic Stress Disorder*, Princeton, NJ: Princeton University Press.

Zita, J.N. (1988) 'The Premenstrual Syndrome: Dis-easing the female cycle', *Hypatia* 3: 77-99.(Arguments for the role of masculine bias in the characterization of PMS and in the social construction of the 'clinical body'.)

Logical atomism

The name 'logical atomism' refers to a network of theses about the parts and structure of the world and the means by which language represents the world. Wittgenstein, in his Tractatus Logico-Philosophicus, expounds a version of logical atomism developed by him around the time of the First World War, as does Russell in works published contemporaneously. It is no accident that their work on logical atomism shares a common surface description since it resulted from their mutual influence at Cambridge. The common theme is that the meaning of our sentences is rooted in a primitive relation between simple expressions and their simple worldly bearers, the logical atoms. In a logically perfect language, atomic sentences describe configurations of these atoms, and complex sentences are combinations of the atomic sentences. But sentences of ordinary language may have a misleading surface form which is revealed as such by analysis. The common theme masks considerable differences of doctrine. In particular, there are differences in the nature of logical atoms and in the arguments for the existence of these atoms.

1 Wittgenstein's logical atomism

Wittgenstein's logical atomism is expounded in his *Tractatus Logico-Philosophicus* (1921) in which he explores the structure of the world and the structure of any language fit to represent the world (see Wittgenstein, L. §5). The unit of sentential representation is the atomic sentence, and Wittgenstein claims that every sentence can be analysed into a combination of atomic sentences which describe atomic states of affairs. Each atomic sentence has exactly one of the two truth-values, true and false, and there are no necessarily true or necessarily false atomic sentences. If an atomic sentence is true, then the state of affairs it describes exists, the existent atomic state of affairs being called 'an atomic fact'. If an atomic sentence is false, then the state of affairs it describes does not exist (see Facts §1; Truth, correspondence theory of).

An atomic sentence is structurally isomorphic to the atomic state of affairs described. It consists of an arrangement of names, and these names have simple objects as their meanings. The atomic state of affairs described by the atomic sentence consists of the named objects arranged in a way which corresponds to the arrangement of the names in the sentence. The objects named in atomic sentences and featuring in atomic states of affairs are Wittgenstein's logical atoms.

The atomic sentences form a system which is fit to represent any possible world. A possible world is represented by an assignment of truth-values to the atomic sentences. Atomic sentences are independent in the sense that there are no logical relations between them which rule out certain assignments of truth-values to the atomic sentences. The actual world is completely described by the atomic sentences which are true and false in it. So the actual world is the totality of atomic facts, the existent states of affairs described by the true atomic sentences.

Atomic sentences form the basis for Wittgenstein's analysis of the sentences of ordinary language. Any sentence of ordinary language has a unique translation into a sentence which is a truth-functional combination of atomic sentences. This is the principle of extensionality. A complex sentence is a truth-functional combination of atomic sentences just in case the truth-value of the complex sentence is entirely determined by the truth-values of its constituent atomic sentences. The sentential connectives of the propositional calculus, such as 'Č&;', yield truth-functional combinations of atomic sentences.

Some truth-functional combinations of atomic sentences yield sentences which are true no matter what truth-values are assigned to the constituent atomic sentences, and sentences which are false no matter what truth-values are assigned to the constituent atomic sentences. These are the tautologies (such as 'P or not-P') and the contradictions (such as 'P & not-P'), respectively. Wittgenstein claims that all necessary truths are tautologies and all necessary falsehoods are contradictions.

Sentences of ordinary language may appear to be atomic, but analysis can reveal them to be complex. The aim of analysis is thus to uncover the accurate representations of the world which lie behind the misleading surface forms of ordinary language. The unit of sentential representation is the atomic sentence, but an atomic sentence is itself compounded out of names which have simple objects as their meanings. Hence, the meaning of every sentence of our language is grounded in this primitive meaning relation holding between names and the simple objects which

they name. But the more familiar entities of ordinary discourse, such as tables and chairs, are complex. So a sentence about a complex entity is bound to be analysed as a combination of atomic sentences, these sentences describing the arrangements of the simple objects which are the parts of the complex entity. Here we have the idea of analysis as the decomposition of complex entities into arrangements of their simple parts.

Why should there be simple objects at all? Why can complex objects not divide for ever and ever into less complex parts? Wittgenstein argues a priori from the possibility of representation to the existence of his simple objects using a *reductio ad absurdum*. First, he assumes that there is an atomic sentence featuring an expression standing for a complex entity. Second, he brings in a background assumption that sense be determinate, that is, every sentence has exactly one of the two truth-values in any possible world. As an illustration, suppose that the sentence 'the broom is to the left of the cupboard' is an atomic sentence. It features the expression 'the broom' which stands for a complex entity. The broom exists only if its parts exist and are arranged in the right way. Consider a possible world in which the broom does not exist - our sentence is not true in this world. But neither can it be false. For the sentence is atomic and thus does not contain as part of its meaning a condition which determines that the sentence is false when the broom does not exist. But every sentence must be either true or false in any possible world. So an atomic sentence must contain only expressions which stand for simple objects.

This shows that simple objects must exist but does not tell us what they are like. It appears that Wittgenstein thought that this was not a logician's job. Early interpreters of the *Tractatus*, taking their cue from Russell's logical atomism and Wittgenstein's examples, took the simples to be the units of perceptual experience, in other words sense-data. But the independence of atomic sentences precludes such an interpretation. 'This sense-datum is red' and 'this (same) sense-datum is green' would be atomic sentences but they cannot both be true.

Wittgenstein later came to think that his model of the relationship between language and the world was radically misconceived. His most immediate criticism was to reject the independence of atomic sentences, which he did by considering sentences featuring colour predicates. This was at the same time the denial that all necessary truths and falsehoods are tautologies and contradictions, respectively. For the necessary falsehood 'this sense-datum is red and green' is not a contradiction. In the first part of his *Philosophical Investigations* (1953) Wittgenstein mounts a sustained attack against the pivotal claims that the meaning of sentences is rooted in the relationship between name and bearer and that the sense of sentences is determinate. Thus the justification for the existence of logical atoms is overturned.

2 Russell's logical atomism

Russell developed his logical atomism over two decades, beginning with 'On Denoting' (1905) and ending with 'Logical Atomism'(1924) (see Russell, B. §11). During this time he continually revised his ideas. (This section sketches the version of logical atomism he expounds in the eight 1918 lectures entitled *The Philosophy of Logical Atomism*.) Russell had originally coined the phrase 'logical atomism' to distinguish his picture of the world as containing many separate things from the opposing picture advocated by the British Idealists, such as F.H. Bradley, according to whom the world is an indivisible whole. His logical atomism justifies his method of analysis, for this method aims to reduce ordinary sentences to complex combinations of atomic sentences, sentences which feature expressions standing for Russell's many separate things.

Wittgenstein and Russell disagree about the structure of atomic sentences. For Russell, atomic sentences consist of simple symbols which divide into proper names and predicates. Simple symbols have simple things (Russell's logical atoms) as their meanings. The mark of a simple symbol is that understanding it consists in being acquainted with its meaning. Russell held that one cannot be acquainted with an ordinary object such as a chair, but only with the perceptual experiences, the sense-data, one has in perceiving the chair, such as the visual experience of a brown rectangular patch and the tactile experience of a rough and hard surface. He also allowed that one can be acquainted with the meanings of predicates, properties and relations. So atomic sentences feature proper names standing for sense-data, and predicates standing for properties and relations of sense-data. For example, 'this is red' is an atomic sentence in which the proper name 'this' stands for a sense-datum and the predicate '…is red' stands for a property ascribed to the sense-datum.

If an atomic sentence is true, it is made true by a corresponding atomic fact which contains the meanings of the simple symbols of the atomic sentence. For example, 'This is red' is made true by the fact consisting of the

sense-datum named by 'this' having the property redness. But Russell's inventory of facts far exceeds Wittgenstein's totality of atomic facts. He also has negative facts which make false sentences false. 'This is red' is made false by the negative fact that the sense-datum is not red. Further, Russell has general facts corresponding to true general sentences such as 'All men are mortal' and 'Some man is mortal'.

Whereas Wittgenstein remains silent about the nature of his simple objects, Russell claims that his simples are sense-data and their properties and relations. This entails that Russell cannot hold that atomic sentences are independent, because two atomic sentences ascribing different colours to a given sense-datum cannot both be true. Russell is more specific about his simples because his argument for their existence turns on an epistemic condition, the principle of acquaintance. This principle holds that the understanding of any sentence consists in acquaintance with the meaning units from which the meaning of the whole sentence is compounded. The principle is satisfied by an atomic sentence because understanding such a sentence consists in acquaintance with the meanings of its simple symbols. A non-atomic sentence must be completely analysed into a sentence which consists entirely of simple symbols standing for objects with which we are acquainted. Analysis thus provides a route whereby the analysed sentence can be understood. The surface form of an ordinary sentence is not a reliable guide to the underlying meaning units which make up the meaning of the sentence. For example, 'Socrates is snub-nosed' seems to feature a proper name, 'Socrates', standing for a person. But Russell claims that the analysis of such a sentence reveals that 'Socrates' is not really a proper name but a descriptive phrase picking out a complicated logical construction from sense-data.

Sense-data and their properties and relations are thus not only the fundamental units of meaning, but also the ultimate constituents of reality. Socrates and his properties and relations are reduced to logical constructions from these ultimate constituents. This reduction has the epistemological advantage embodied in Ockham's razor. By taking Socrates to be a logical construction from sense-data, one reduces the risk of error because beliefs about sense-data and constructions from sense-data are supposed to be more secure than beliefs about objects which stand behind the sense-data.

Russell's logical atomism was well-received by the logical positivists who confused it with Wittgenstein's logical atomism (see Logical positivism §3). They fixed on Russell's reductionist programme for translating empirical sentences into sentences about sense-data. This programme is fraught with difficulties both of principle and detail. One chief difficulty, promoted by Wittgenstein in his *Philosophical Investigations*, is the alleged incoherence of a private language consisting of reports of a single person's sense-data (see Private language argument).

See also: Knowledge by acquaintance and description; Pluralism

ALEX OLIVER

References and further reading

Anscombe, G.E.M. (1959) *An Introduction to Wittgenstein's Tractatus*, London: Hutchinson. (Chapter 1 argues against the view that Wittgenstein's logical atoms are sense-data.)

Griffin, J.P. (1964) *Wittgenstein's Logical Atomism*, Oxford: Clarendon Press.(A clear account of Wittgenstein's logical atomism, emphasizing the influence of Hertz's ideas about models in natural science.)

Pears, D.F. (1967) *Bertrand Russell & the British Tradition in Philosophy*, London: Fontana. (A detailed exposition of the development of Russell's logical atomism.)

Russell, B. (1905) 'On Denoting', in *Logic and Knowledge*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 41-56.(Introduces the Russellian theory of definite descriptions which lies at the heart of Russell's logical atomism.)

Russell, B. (1911) 'Knowledge by Acquaintance and Knowledge by Description', in *Mysticism and Logic*, London: Longman, 1918, 209-32.(States the principle of acquaintance.)

Russell, B. (1914) *Our Knowledge of the External World*, London: Allen & Unwin.(Chapters 1-4 contain the rudiments of the metaphysics of facts and logical constructions from sense-data.)

Russell, B. (1918) 'The Philosophy of Logical Atomism', in *Logic and Knowledge*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 177-281.(The version of Russell's logical atomism sketched in this entry.)

Russell, B. (1924) 'Logical Atomism', in *Logic and Knowledge*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 323-43.(A review of the principles of Russell's logical atomism in which sense-data drop out of the picture.)

Urmson, J.O. (1956) *Philosophical Analysis*, Oxford: Clarendon Press.(An example of the conflation of Russell's and Wittgenstein's logical atomism, but useful for tracing the influence of logical atomism on the

logical positivists.)

Wittgenstein, L. (1921) *Tractatus Logico-Philosophicus*, trans. D.F. Pears and B.F. McGuinness, London: Routledge & Kegan Paul, 1974.(Wittgenstein's statement of his logical atomism.)

Wittgenstein, L. (1929) 'Some Remarks on Logical Form', supplement to *Proceedings of the Aristotelian Society* 9: 162-71.(Rejection of the independence of atomic sentences.)

Wittgenstein, L. (1953) *Philosophical Investigations*, trans. G.E.M. Anscombe, Oxford: Blackwell. (Paragraphs 1-242 undermine the fundamental principles of 'the author of the *Tractatus*'.)

Mass terms

Mass terms are words and phrases such as 'water', 'wood' and 'white wallpaper'. They are contrasted with count terms such as 'woman', 'word' and 'wild wildebeest'. Intuitively, mass terms refer to 'stuff'; count terms refer to 'objects'. Mass terms allow for measurement ('three kilos of wood', 'much water'); count terms allow for counting, quantifying and individuating ('three women', 'each word', 'that wildebeest over there').

Philosophical problems associated with mass terms include (1) distinguishing mass from count terms, (2) describing the semantics of sentences employing mass terms, and (3) explicating the ontology presupposed by our use of mass versus count terms. Associated with these philosophical issues - especially the third - are the meta-philosophical issues concerning the extent to which any investigation into the linguistic practices of speakers of a language can be used as evidence for how those speakers view 'reality'.

1 Distinguishing mass and count terms

The distinction between mass terms ('water', 'wood' and so on) and count terms ('woman', 'word' and so on) can be seen as syntactic, semantic or pragmatic. If the distinction is seen as syntactic, one might remark that mass terms occur with the quantifiers 'much' and 'little' and with the unstressed article 'some', that they are susceptible to measurement phrases such as 'litres of' and 'amount of', and that they do not exhibit a singular/plural distinction. Conversely, count expressions occur with the quantifiers 'each', 'every', 'many', 'several', 'few' and the stressed 'some', use the indefinite article 'a(n)', are susceptible to counting phrases such as 'five' and 'a score of', and exhibit a singular/plural dichotomy manifested in the count term itself and in agreement with the verb phrase.

If it is seen as semantic - a distinction between the different ways that mass and count terms refer - then one might remark that count expressions refer to discrete, well-delineated objects while mass terms refer without making it explicit how the referent is individuated (some have said that the referents of mass terms are continuous rather than discrete). This feature of mass reference gives rise to the 'cumulative reference test' (any sum of parts which are M is also M) and to the 'distributive [homogeneous] reference test' (any part of something which is M is also M).

If the mass/count distinction is seen as a pragmatic one, then one will look to how people use count terms to 'individuate' the world. This gives rise to such tests as whether there is a definite answer to the question 'How many X's are there in such-and-such place?'. In the philosophical literature (following Strawson 1959), terms which pass this counting test are often called 'sortal terms' - although they are equally often called count terms - and ones that fail the test are called mass terms. In this literature, with its emphasis on the pragmatic notions of 'identifying' and 'individuating', it is common to deny that such terms as 'thing', 'object' and so on are sortal (count), there being no definite answer to the question 'How many things are in the room?'. This is so despite the fact that such terms clearly satisfy the syntactic criteria.

All these tests - the syntactic, the semantic and the pragmatic ones - have been challenged. Writers have pointed out that mass terms such as 'wood' can also be used as count: 'a wood' might designate oak or spruce, for example. And 'wildebeest' can be used as a mass term: 'He's not really a vegetarian; he eats wildebeest'. Furthermore, a universal grinder would take an object that an alleged count term referred to - a chair, for example - and grind it up into a powder so that then there would be chair all over the floor. (This last sentence uses 'chair' in a mass manner, thereby showing that the language already has this usage of any count term in the background. And this usage exists despite the fact that we could also have said 'There is wood all over the floor'.)

The criteria just mentioned are usually seen as applying to entire noun phrases as well as to the simple nouns themselves. Furthermore, some writers attempt to apply the mass/count distinction to other syntactic categories, especially adjectives, verbs and adverbs. The application to verbs is especially interesting (see Mourelatos 1978).

2 Semantics and ontology of mass terms

The problem with giving a formal semantic analysis of mass terms arises because first-order predicate logic appears to assume that the entities in the domain of quantification are individuals, so it only makes sense to characterize them with count nouns. When we say, in the quantifier idiom, 'For all x, if x is F…', it is apparently

assumed that the items in the domain have already been individuated. For if *F* were to be interpreted as 'snow', for example, what would be the values of *x*?

Famously, Quine (1960) held that mass terms are ambiguous: when in 'subject position' they are singular terms (names), but when in 'predicate position' they are general terms (predicates) which are 'true of each portion of the stuff in question, excluding only the parts too small to count'. As a name (when in subject position), Quine holds that a mass term 'differs none from such singular terms as "mama"…, unless the scattered stuff that it names be denied the status of a single sprawling object'.

This proposal has not satisfied various authors, who have objected to the nonuniform treatment and to various logical consequences of this approach. For example, on Quine's analysis, 'Water is wet, and this puddle is water' does not imply 'This puddle is wet'; and 'Water is water' does not come out a logical truth. Writers after Quine have proposed many different approaches. Possibly the most popular alternatives involve mereology, according to which the main operator is 'is a part of'. Mass (and other) terms are taken to designate 'mereological wholes'. Some authors have grafted onto pure mereology a notion of 'having certain structural properties', so as to avoid the minimal parts problem alluded to in Quine. (The atoms, *inter alia*, which are part of water are 'too small' to count as water.) But these theories also have not satisfied all those involved in this area, usually because the treatment of certain logical inferences is thought incorrect: the formal semantic analyses do not mirror intuitive beliefs concerning logical consequence.

An alternative is to retain the idea that mass terms name some kind of object - a 'substance' - and to invoke a relational predicate such as 'is constituted of'. This presents a number of tricky issues and there are types of sentences for which such an analysis is not obviously suitable, but still various authors have adopted it. Besides the formal differences entailed by these two approaches (mereological calculus of individuals versus classical logic with a relational constant of constitution), there is an ontological difference, for mereological wholes are generally taken to be physical whereas substances or kinds are often viewed as abstract entities (see Substance).

Another formal semantics of mass terms invokes sets as their denotation. Differences among theorists can then be seen as differences about what the sets contain. One question on which theorists differ is whether the sets contain only 'minimal entities' - the smallest items to which the mass term refers (flakes, maybe, for 'snow'; the items and size vary according to the mass term in question) - or whether it should contain 'ordinary objects' (flakes, drifts, snowmen, snowballs and so on; any object which can be said to be snow). The former proposal has not gained many adherents due to the difficulty of specifying a set of 'minimal entities' for such mass terms as 'garbage', 'speed' and 'information'. The latter proposal runs into difficulties in trying to account for the denotation of definite noun phrases (NPs) such as 'the snow on the table'. It is generally not true that there is exactly one snow-thing on the table. (There is one ball and also many flakes making it up, for example.) So the only reasonable proposal is for the NP to designate *all* the snow-things on the table. But then certain measurement sentences - for example, 'The snow on the table weighs one kilo' - come out wrong, since we will count the same snow-entities many times over.

Theories of mass terms show a fundamental division between those that are committed to abstract substances and those that are physicalistic in nature, invoking mereological wholes. On the physicalistic side are those theories which propose that the ontologically basic objects are the minimal entities, those which claim that the larger entities are 'constructions' out of these minimal entities, and those theories which propose that all these entities are equally basic. On the other side of the gulf are the various styles of substance theories, which usually invoke a lattice structure of kinds. Such ontological issues are discussed in Pelletier and Schubert (1989) and in Burkhardt and Smith (1991).

See also: Logical and mathematical terms, glossary of; Mereology

<div align="right">JEFFRY PELLETIER</div>

References and further reading

Bunt, H. (1985) *Mass Terms and Model-Theoretic Semantics*, Cambridge: Cambridge University Press.(This is the longest and possibly most thorough work on mass terms. It is aimed at a computer implementation of a natural language understanding system which will include mass terms. It invokes 'ensemble theory', which is a type of atomic mereology.)

Burkhardt, H. and Smith, B. (1991) 'Mass Terms', in *Handbook of Metaphysics and Ontology*, Munich: Philosophia.(This entry goes into the metaphysical and ontological presuppositions of mass terms in more detail than is possible here. It includes some speculation about how the mass/count distinction might have arisen in natural language.)

Lønning, J.-T. (1987) 'Mass Terms and Quantification', *Linguistics and Philosophy* 10: 1-52.(An example of the direction that research into the mereological interpretation of mass terms has taken.)

Mourelatos, A. (1978) 'Events, Processes, and States', *Linguistics and Philosophy* 2: 415-34.(Mourelatos draws explicit attention to the relationships that hold between certain types of verbs and verb phrases and the mass/count distinction.)

Pelletier, F.J. (ed.) (1979) *Mass Terms: Some Philosophical Problems*, Dordrecht: Reidel.(This anthology includes many of the classic articles on the topic of mass terms, especially in the philosophical tradition. It also includes a comprehensive bibliography up to 1978.)

Pelletier, F.J. and Schubert, L.K. (1989) 'Mass Expressions', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic*, Dordrecht: Reidel, vol. 4, 327-407, esp. 391-4.(The most thorough summary of the work done on mass terms, both linguistic and philosophical. It also proposes two theories to account for the phenomena, and includes a comprehensive bibliography of work from 1978 to 1988, plus earlier work where relevant.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press, 91-100. (Although this book has only a small section on mass terms, it formed the starting point of most future formal work on the topic. Quine proposed his 'dual analysis' here and asserted that mass terms come from an 'early, primitive stage in linguistic evolution'.)

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985) *A Comprehensive Grammar of Contemporary English*, London: Longman.(The most thorough descriptive account of the grammar of the English language, with a section about the mass/count distinction, especially as it is used in normal speech. It does not deal particularly with the formal semantics of the distinction.)

Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen.(It is here that the distinction between 'feature-placing universal' and 'sortal universal' was drawn. Many writers wished to put mass terms in the former category, giving rise in the philosophical literature to the opposition 'mass/sortal'. Other writers thought of mass terms as a type of sortal term, while still others argued for a third category.)

Meaning and communication

The two fundamental facts about language are that we use it to mean things and we use it to communicate. So the philosophy of language tries to explain what it is for words and sentences to mean things and also what it is for us to communicate by using them. Although it cannot be accidental that meaning and communication go together, it is quite easy to see them as fundamentally distinct. Thus on some accounts the meaning of sentences is conceived in terms of a 'representative' power whereby they stand for either aspects of the world or ideas in the mind and their use in communication is derived from this property: language serves as a vehicle for meaning, itself thought of in independent terms. An alternative approach seeks to link the two more closely, seeing representation as itself only possible through the use of terms of a common language, used in communication.

At its most primitive, communication may be simply akin to infection, as when one animal communicates fear or hostility to another. Simple signalling is also described as communication, as when bees communicate the direction and quantity of pollen to other bees, or when birds communicate territorial or sexual claims. Here the signaller issues a sign and the recipient modifies its behaviour upon perceiving it; the content of the signal is interpreted (by us) to be whatever goal seems to be served by issuing it: maximizing the success of the hive in obtaining honey, or keeping other birds out of a particular space, for example. Humans communicate in a similar sense, through the use of emotional and other signals, unconscious 'body language' or equally unconscious or semiconscious signs, which may include fashion statements or indications of status, for example. When we communicate successfully we share an understanding. Everything is open between us; nothing is concealed nor taken by one person in one way and another in a different way. With linguistic communication this common understanding can be put in terms of meaning: we each know what the other means. This implies that we understand not only the semantics of the utterance, but also the pragmatics: if the utterance states that something is the case then we know what this is, but we also know if the statement was meant ironically, or condescendingly, or with some other intent (see Pragmatics). It is plausible to think that the basic goal of speech is to achieve such communication, although there will be cases where we speak in order to conceal our thoughts, or to mislead the hearer. People fluent in the same language achieve open communication more or less effortlessly, at least where straightforward messages are concerned. But philosophers of language have not found this phenomenon easy to understand. Issues arising include: the place of intention in communication (see Communication and intention); the place of linguistic convention in communication (see Language, conventionality of); the relationship between shared understanding and activities such as translation and interpretation (see Radical translation and radical interpretation); the pervasive possibility of indeterminacy; and the nature of rules and the extent to which language is essentially social - and indeed the extent to which language is necessary to communication (see Meaning and rule-following). In the analytic tradition, major writings on this subject include works by Wittgenstein, Grice, Lewis, Bennett, Kripke, Searle and Tuomela, while a more continental interest in the field is especially illustrated by Habermas (1984, 1987).

The classical tradition, including Aristotle, Hobbes and Locke, thinks of communication in terms of one party having an idea then using language as a medium for transmitting it to others. The meaning of the linguistic vehicle is then derived from the primitive representative powers of the idea. The analytic tradition has sought to bring our ability to represent things closer to our ability to express what we think linguistically, even to the point of reversing the dependency and making thought itself (or at least the kind of thought that could be expressed in language, unlike perhaps the thought of the graphic artist or the musician) dependent upon its linguistic expression. The question then becomes one of saying how language can generate 'intentionality', which is the power of being 'about' external and often absent situations which it represents truly or falsely (see Intentionality). If, for example, language is essentially a shared and social construction, and if it is necessary to thought, then we derive the result that a born Robinson Crusoe would not be capable of thought, at least until he acquired a social identity. The consequence seems uncomfortable, since it is easy to imagine such a Crusoe giving all sorts of signs of intelligent adaptation to his environment, and conceiving and executing complex plans. For some thinkers, such as John Searle (1983), this is because intentionality resides in our biological nature, in advance of and independent of language. For others, such as Jerry Fodor (1987), it is because we come equipped with an innate representative medium conceived on the model of a natural (social) language, but itself explaining the powers of natural languages: a 'language of thought' (see Language of thought).

Arguably too much of the theory of communication treats the process as one of interpretation or translation, in which the task of the hearer is to theorize about the thoughts intended by some act of communication of the speaker. This makes it seem as if each person is secure in their own 'ideolect', with it being relatively problematic whether the same ideolect is shared by others. This clearly distorts what it is to share a language, which surely entails finding ourselves (surprisingly literally) of one mind about meaning. We have a securely shared common understanding, in which I have no privileged access to my own meanings which is not communicable to others. Philosophers have therefore struggled against the vision of the infant, inducted into this shared language, as a kind of outsider or 'little linguist' busy theorizing (in what medium?) about the potential meanings of the utterances made by surrounding people. The proposed alternative derives from the *Verstehen* (literally 'to understand') tradition of humane sciences (see Dilthey, W.), seeing the task of the infant not so much as one of theorizing as one of imitating and simulating the expressions of others. I achieve the openness of full communication with you, on such an account, when I know what it would be to make your words my own.

SIMON BLACKBURN

References and further reading

Bennett, J. (1976) *Linguistic Behaviour*, Cambridge: Cambridge University Press.(A lively development of Grice's work, especially valuable for its discussion of the relation between full intentionality and more primitive signalling systems.)

Blackburn, S. (1984) *Spreading the Word*, Oxford: Oxford University Press.(An attempt at an accessible discussion of most of the central issues, aimed at moderately advanced students.)

Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.(The classic defence of the theory that meaning is the outcome of computational structures in the mind/brain.)

Grice, H.P. (1957) 'Meaning', *Philosophical Review* 66; repr. in *Studies in the Way of Words*, Cambridge, MA, and London: Harvard University Press, 1989.(The seminal article in the field.)

Habermas, J. (1984, 1987) *The Theory of Communicative Action*, trans. T. McCarthy, Boston, MA: Beacon Press, 2 vols.(An attempt to connect the social nature of language to the virtues of free democratic debate.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell. (A particularly accessible presentation of the thesis that meaning is essentially social.)

Lewis, D.K. (1969) *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.(The classic modern analysis of the concept.)

Searle, J.R. (1969) *Speech Acts*, Cambridge: Cambridge University Press.(An accessible development of the work of J.L. Austin on the actions that generate meaning.)

Searle, J.R. (1983) *Intentionality: An essay in the philosophy of mind*, Cambridge: Cambridge University Press. (Searle's work is perhaps the most impressive development of the notion of acts performed in speech, but also controversial through its firm belief that the intentional powers of language are derived from a prior biologically engendered capacity.)

Tuomela, R. (1984) *A Theory of Social Action*, Dordrecht: Reidel.(An influential attempt to see collective action as arising from the individual intentions of the different parties.)

Meaning and rule-following

Wittgenstein's discussion of rules and rule-following, and the recent responses to it, have been widely regarded as providing the deepest and most challenging issues surrounding the notions of meaning, understanding and intention - central notions in the philosophy of language and mind. The fundamental issue is what it is for words to have meaning, and for speakers to use words in accordance with their meanings. In Philosophical Investigations and Remarks on the Foundations of Mathematics, Wittgenstein explores the idea that what could give a word its meaning is a rule for its use, and that to be a competent speaker is to use words in accordance with these rules. His discussion of the nature of rules and rule-following has been highly influential, although there is no general agreement about his conclusions and final position. The view that there is no objectivity to an individual's attempt to follow a rule in isolation provides one strand of Wittgenstein's argument against the possibility of a private language.

To some commentators, Wittgenstein's discussion only leads to the sceptical conclusion that there are no rules to be followed and so no facts about what words mean. Others have seen him as showing why certain models of what it takes for an individual to follow a rule are inadequate and must be replaced by an appeal to a communal linguistic practice.

1 Meaning and rules

Words are meaningful only if there is such a thing as using them correctly or incorrectly. To have determinate meanings they must have application in some situations and not others. Words that had application to just anything would be robbed of meaning and could not be distinguished from one another. But what gives a word meaning? And how does this determine what uses do and do not comply with it? Answers to these questions must be compatible with a credible epistemology of meaning: when we understand a word we must be able to use it in accordance with its meaning. If something were to settle the right or wrong use of a term, we must be able to take cognizance of this. The objectivity of meaning and linguistic judgment depends on this. We have to secure both of these simultaneously.

In his later writings, Wittgenstein rejects the idea of meanings as mental or abstract entities to be associated with particular signs. Instead he takes the meaning of a sign, or word, to be its use in a language. However, in equating meaning with use he does not assume that every way of using a sign can contribute to its meaning. To have a meaning there must be some particular range of application that counts as using a word correctly. But what does the correct use of a word consist in? Wittgenstein explores the claim that a word is used correctly when it is used in accordance with a rule. If there are rules governing the use of words, then there are standards speakers have to meet to deploy words competently. Competence requires speakers to know what counts as applying the word correctly or incorrectly on any occasion. The appeal to rules is also thought to guarantee the range of things a word applies to: the rule can be thought of as settling the application of a term not only to cases considered so far, but also to hitherto undiscovered cases.

These features of rules bring out the normative element in meaning. Given that words have meanings there is something that counts as using them in accordance with these meanings: this is how words *should* be used. When meanings are given by rules of use, competent speakers are required to conform to these rules: a rule does not just describe the use we make of an expression, it says how we *ought* to use it.

To satisfy ourselves that there is some substance to the notion of meaning, we need to account for the nature of these rules and the requirements they impose on our correct use of words. We also need an account of how speakers succeed, if they do, in following these rules. The issues are connected. If no account can be given of what it takes to heed the requirements of a rule, it is hard to credit those rules with any normative force in determining the meanings of words in our language. On the other hand, the idea of us conforming our linguistic practice to the requirements of rules cannot be maintained if we can give no substance to the idea of there being rules to be followed in the first place.

2 Rules and rule-followers

Wittgenstein addresses the issue of what it is to follow one rule rather than another by considering the case of a

learner asked to continue the arithmetical series: 2, 4, 6, 8…. There are infinitely many options for them. They could go on by writing 10, 12, 14, or 10, 14, 18, and so on. All they are shown are finitely many cases and from there they must acquire the ability to continue the series the mathematician intended. To go on from any position to the next they need to know the rule for expanding the series. But how do they acquire that from their exposure to the examples? Wittgenstein reviews a number of models for such knowledge of rules, and rejects them as inadequate.

First, we might think of learners giving themselves an instruction (such as $n = x + 2$), which tells them what to do next. But no instruction is sufficient, since it can only be followed if the learners know what it means. If they know this, then we must ask in what their acting in accordance with the meanings or rules governing *that* instruction consists. Either we appeal to further instructions, thus setting up an infinite regress, or else we suppose they just know what the instruction means in which case we have no explanation of their understanding, and we are no further forward.

Perhaps a learner simply reacts to the presented cases in some way, and either does or does not satisfy the mathematician. But if this learner is simply caused to respond, we are unable to say what it is for them to go by the rule as opposed to merely reacting to instances of it. On a future occasion they could be caused to act in a way at odds with the intended series. Also, their reactions may be fortuitous: they may be fitting the rule but not following it, blind to what is required of them. Only if the learner and the mathematician follow the same rule is the possibility of coincidence neutralized.

The justification of each move, if there is one, will depend on the rule the learner is following. But which rule is this? There are infinitely many rules compatible with the learner's behaviour so far, and every move they subsequently make can be made to accord with some rule or other. So what reason is there to think, before the learner decides on their next move, that there is some particular rule they are following? Perhaps we do not consult anything to decide how we are using a sign (or expanding a series). Nevertheless, the temptation exists to say that there must be something that determines whether we are going right or wrong in what we decide to do.

We might then say the rule itself determines right and wrong moves, because it consists in the full expansion of the series with all the instances fixed. As Wittgenstein puts it, it is as if all the steps had already been taken. This is a picture of rules as rails stretching ahead of us to infinity, with the rules fixed in some abstract or Platonic realm: what is required of us in any given application of them would thus be constituted wholly independently of human propensities to judge. The difficulty with this picture is that it gives us no idea of how we recognize which standards to conform to. If rules were blankly external to the mind, there would be no way to discover what they require of us. However we used words, whether or not we coincided with these objective standards would be a matter settled independently of us. Such a conception makes it unclear how such rules or standards could inform our practices, and it is hard to see how we could attempt to conform to rules that fall outside our cognitive reach.

The Platonist's picture mislocates the normative dimension of speech. At best our judicious use of words would be a series of stabs in the dark with which we hoped to speak significantly by coinciding with the mind-independent meaning-facts. This picture may afford the idea of an objective standard for our linguistic judgments, but there would be no difference (from the point of view of the speaker) between observing this standard and following one's subjective inclination to use the words as one felt fit. Without speakers being apprised of the standards of use that regulate their linguistic practice, meaning-scepticism would arise, putting in doubt both the existence of Platonized rules and the possibility of a speaker using words in accordance with their meanings. It seems as if the trouble with Platonism is that it locates the meaning-determining rules wholly outside the mind of the speaker.

Wittgenstein next turns to the temptation to find some self-interpreting item in the mind which serves to stop the regress of interpretations, and by itself points to the correct application of a word. Wittgenstein tells us that when we review the instructions we might give ourselves for deploying a word, it is as if meaning was the last interpretation. The trouble now is that there are no candidates for this role. If we suppose that the meaning itself is an item in mind, or that we have an intuition of what is required of us on any occasion of use, we need to know how such things as items in consciousness could point to the conditions of application for a word. For if intuition leads us to apply the word in one way, it could just as easily mislead us: we are simply reacting in a way that feels right. No items whose properties we can scrutinize consciously seem to carry the right information about how we should use a word, or comply with a rule in unexamined cases. It is utterly mysterious how anything we

consciously consult could do this. It might be said that knowledge of our intention in using the word will serve here, but this also proves unsatisfactory: we are trying to explain how we can have such an intention in the first place, and how we can tell whether or not we are conforming to it. The appeal to an inner item is thus no better than Platonism. Locating the item inside the mind has still not connected our use of a word to a norm we can appreciate and whose requirements we could take into account in our linguistic use.

We appear to have exhausted the options. Does this mean that when we judge how to use an expression there is nothing that counts as getting it right? To concede this would be to surrender the objectivity of linguistic judgments and settle for an absence of genuine constraints on how to use words. However we go on using a word in future, there would be nothing that counted as acting in accordance with the rules, since there would be nothing for the requirements of the rules to consist in. How could thinking about a rule show me what I was meant to do at this point? There may be a subjective impression of constraint, of going on in the right way. But is there anything this answers to? For any way I chose to use the word in a new case, can be made to conform to some rule or other. There are interpretations that can be made consistent with anything I do, as Wittgenstein tells us in *Philosophical Investigations*: 'any interpretation hangs in the air along with what it interprets, and cannot give it any support. Interpretations by themselves do not determine meanings' (1953: §198). If every way of employing a term can be made to conform to a rule, then 'whatever is going to seem right to me is right. And that only means that here we cannot talk about "right"' (1953: §258). Without a distinction between what 'seems right to me' and what *is* right in my use of language, there can be no objectivity to meaning.

This is the conclusion Saul Kripke sees Wittgenstein as reaching at §201 of *Philosophical Investigations*; a conclusion Kripke, in his influential reading of Wittgenstein, calls the sceptical paradox.

3 Kripke's Wittgenstein

As Kripke (1982) formulates it, the sceptical paradox arises for a subject who has already been using the symbol '+' satisfactorily to compute sums of numbers less than 57. Kripke's example supposes that they are then asked to add 68 and 57. They might answer '125', but what fact about them makes this the right answer? The challenge to the subject is to come up with a fact about previous mental, behavioural or physical history that constitutes meaning *plus* rather than the deviant operation *quus*: an operation which shadows addition until we reach numbers greater than 57 at which point the answers are not the sum of those numbers, but always 5. If they meant *plus* by '+', then they are right; if previously they meant *quus*, then to answer correctly they should have answered '5'. They are only justified in answering '125' if there is a fact about what they mean by '+', a fact about which rule they have been following until now and continue to follow.

For Kripke this is not an epistemological problem about there being certain facts to which we cannot have access. He allows the sceptic to suppose that a subject has perfect recall and access to their past behaviour, conscious experiences, memories and so on, but points out that they are still unable to come up with any fact that constitutes their meaning *plus* rather than *quus*. The failure to do so leads Kripke's sceptic to conclude that there is no fact of the matter about a subject's meaning one thing rather than another. There are simply no facts of the matter concerning meaning. (Though Simon Blackburn (1992) argues that this claim cannot be made in the first person without it being self-defeating since the person must be able to distinguish *plus* from *quus* to appreciate the problem.) Although Kripke takes this damaging conclusion to be the upshot of Wittgenstein's rule-following considerations, he suggests that Wittgenstein tries to show that we can live with the consequences. (For Kripke's presentation of what he takes to be Wittgenstein's sceptical solution, see §4.)

Crispin Wright (1989a) and John McDowell (1984) both challenge Kripke's reading of Wittgenstein. They point out that Kripke's sceptic is forcing us to provide a reductionist answer to the question of what my meaning *plus* rather than *quus* consists in. To substantiate the claim to mean one thing rather than another, we are asked to come up with some fact other than that we mean *plus*, or had the intention to use the word as we previously did. McDowell suggests there is no way to avoid an appeal to the fact that we meant *plus* by our words all along, and he insists that we have to take this as a stopping place for explanation. No justification of this fact can be given: we simply have to remind ourselves of the meanings we know our words to have. This approach has come to be known as quietism (see §5). Wright thinks we have to appeal to speakers' previous intentions concerning their use of the word to settle questions of justification, but unlike McDowell he believes there is a constructive explanation of what it is to know one's intentions and conform to them (see §6).

Wittgenstein's remarks at §201 of the *Philosophical Investigations* certainly suggest a sceptical paradox which gives rise to the dilemma McDowell identifies: either we have an item to provide the interpretation and are saddled with Platonism or a regress of interpretations, or in using words as I do I simply act blindly and there is no fact of the matter about what I mean. According to McDowell, the paradox depends on an assumption to be discharged: namely that meaning something by a word must be a matter of attempting to give it an interpretation. In the second half of §201, Wittgenstein goes on to suggest that there must be a way of meaning something that is not a matter of interpretation. This recommends a different way of construing the presence of rules in linguistic behaviour. In following a rule, the individual does not search for an interpretation. He just acts without giving himself supporting reasons. But unlike Kripke's reading, this does not lead to the conclusion that there is no fact of the matter about what a speaker means. We are told at §289, that 'To use an expression without justification is not to use it without right.' What is needed is an alternative account of what it is to act in accordance with a rule and to use our words correctly.

This is where the appeal to the community comes in. Rather than train our sights on facts about an individual, following a rule is a matter of participating in a communal practice. This retains the idea that there must be something that counts as acting in accordance with a rule or going against it on any occasion, but now conceives of it as a matter of obeying a practice or going against it. The idea of rules as consisting in communal practices leaves room for a number of different treatments of the way an appeal to the community can help here. These are reviewed in §4 as attempted solutions to the sceptical paradox, after discussion of another popular response to Kripke: the dispositional response.

4 Attempted solutions

Dispositional response. The dispositionalist attempts to respond to Kripke's sceptic by insisting that while no occurrent facts about a speaker's behaviour or mental history determine what they meant by their words, their disposition to respond in particular ways to cases when confronted by them settles the matter. The dispositionalist is here relying on the fact that when a thinker is asked what ' $68 + 57$ ' is, they will be disposed to reply '125'. However, the disposition of the subject can be tested in only finitely many cases, whereas addition is defined over an infinite range. Also, the subject may not have the disposition to respond in the case of very large sums where the computations are difficult, and they may be disposed to make mistakes. These cases could be dealt with by pointing to a different set of dispositions to work out the sum on paper and to check the results. However, Kripke's main objection to the dispositionalist's solution is that it loses sight of the normative element of meaning. In each case, the dispositionalist tells us how we *would* respond but not how we *should* respond. If I am using a word in accordance with its meaning, there is a way in which I am obliged to use it if I want to apply it correctly. Kripke tells us that the dispositionalist simply equates correctness and performance.

Appeal to the community. One can perhaps resist the conclusion that there is no fact about what a speaker means by their words by bringing in a background of other users against which the use of a word can be measured. We could say that the speaker does attach a certain meaning to a word, if their use of it coincides with the way the majority of people in their linguistic community use it. This response parallels the dispositional response, except that it considers the dispositions of everyone in the linguistic community. The notion of correctness is established in this case not by appeal to a rule that the individual is following, but by appeal to a widespread pattern of use exemplified by fellow language-users. In so far as there is a fairly widespread and consistent pattern of use for a word, the individual will count as using a word correctly just in such cases as their use of the term coincides with this pattern. They need not know that it coincides, or even check that it does. On this view, a speaker's knowing what they mean will not be the effortless and direct matter it is for most people. For this speaker it will be a matter of determining their fit with the uses other people make of these words. We may submit to correction when a number of people point out how our use of a word diverges from theirs, but we can be offered no determinate advice on what is required of us in the future if we are to continue conforming to majority use. For at the community level, there is no fact of the matter about what we mean collectively or what rules we are following (see Wright 1980). The strategy establishes for each individual language-user a sense of their using words correctly or incorrectly, and to that extent determines some objectivity for their linguistic judgments, but the community itself does not go right or wrong, it simply goes. Such a strategy puts in doubt the existence of meaning for shared public languages. And as far as individuals are concerned, it does nothing to guide their use of language.

They will merely coincide or fail to coincide with their fellow language-users.

Kripke's sceptical solution. Kripke's sceptical solution invokes a different appeal to the community. Unlike a straight solution this in effect concedes that there is no answer to the sceptical paradox, but attempts to show us that we can live with the consequences. Kripke maintains that although there are no meaning-facts to settle how my words *should* be used, I will go on using them like other people and be subject to correction by them. Claims about what my words mean will lack truth-conditions but, Kripke insists, they will have assertion-conditions: conditions under which people will be prepared to assert that someone is using a word in conformity with its meaning. There will be nothing about someone's use of a word to which this claim is answerable, but there will be uses we have for such statements that do not require their truth. There is no such thing as individuals taking themselves to mean something in isolation from the community, since there would be no background consensus to establish even assertion-conditions for claims about what they mean. The consensus depends on a community, so the appeal to the community rules out the possibility of someone using a private language.

The sceptical conclusion has been resisted by those who maintain that it is part of Wittgenstein's aim to defuse rather than accept the sceptical paradox. The paradox arises, it is claimed, because of a mistaken understanding of what is required to follow rules competently.

5 McDowell's quietism

As long as there is something that counts as obeying a practice or going against it on any occasion, there is such a thing as correctly following a rule. Rules will no longer be thought of a something I consult inwardly, or by reference to which I interpret what is required of me. Rather, they will simply be enshrined in our practices - the normative practices which inform our linguistic behaviour and impose constraints on the meaningful use of our terms. Individuals follow rules when they participate in communal practices. It has to be by reference to this notion of practice, that our judgments about how to use words come to have objectivity. So what is a practice and how does it provide for the idea of going right or wrong on any occasion?

The simple idea of *de facto* conformity with others cannot serve here. For according to that conception, being right in our use of a word is just a matter of whether we find ourselves in agreement with the majority, and this threatens to deprive the picture of its normative dimension (see §4). How can whether I am in step with my community determine how I should go on using and understanding a term? How can this show me how I ought to use language to be faithful to the meanings of my words?

What we need is to restore the idea of a community of speakers who operate according to the norms of meaning. John McDowell (1984) insists that such a picture is available to us when we observe Wittgenstein's warning not to try to dig below bedrock (the level, Wittgenstein says, at which justifications give out). We can dig down to the ground to find justifications but we will reach a level where no further explanation can be given by reference to what goes on at the level below. This is bedrock. McDowell characterizes bedrock for our linguistic practices as a level of activity at which norms are still in place, and where we must describe one another's linguistic behaviour in the meaningful terms available to us as speakers of the language in question. We will have to describe people's activities as 'saying that *such and such*' - where the words we use to describe what is said (*such and such*) are the very words used by the speaker. This kind of description will not be intelligible to those outside the practice.

This is a resolutely anti-reductionist approach to meaning and rule-following, and we are told that we can expect no explanations of meaningful speech and linguistic practice by reference to something else. We simply have to remind ourselves how familiar we are with the meanings of our words and what they require of us in so far as we participate in these normative practices. No explanation can be given of their significance, and none is called for. This position known is as quietism.

6 Wright's extension-determining approach

Crispin Wright (1989a) opposes both quietism and the sceptical paradox, believing that there is room to show what the rule-governed meaning-facts consist in. Wright accepts the challenge to show how it can be that the requirements of rules governing our words are not simply up to us, while it is also true that they are not constituted wholly independently of us. The aim is to avoid surrendering the objectivity of meaning while at the same time bringing it within epistemic reach.

Wright argues that what constitutes a correct application of a word, including its application to a new case, and how it was used previously, is not something that can be settled independently of human judgment. A contribution from us is required to settle the matter. Wright avoids a slide to subjectivism by claiming that our best opinions about what we previously meant or intended will settle the matter, and since not every opinion we form is best it is not true that whatever we judge to be right is right. Only those judgments we make in the most propitious conditions for judging go into determining the right use of a term. The key point is that our linguistic judgments under these conditions *determine* rather than *reflect* the correct application of our terms. These judgments about how to use words do not involve a tracking epistemology: we are not keeping up with Platonic standards or internalized instructions. According to Wright, if we can specify cognitively ideal circumstances for judging in a substantial and non-circular way, then there is nothing but our opinions in those circumstances to decide what we mean by these words. In this conception, meanings are like secondary qualities. Objections to this account question the plausibility of giving a non-circular specification of the ideal circumstances for judging meanings or conditions of application.

See also: Private states and language; Language, social nature of

BARRY C. SMITH

References and further reading

Blackburn, S. (1992) 'Theory, Observation and Drama', in *Mind and Language* 7: 187-203.(Aims to show how appeal to the first-person case can defeat Kripke's sceptic. Referred to in §3.)

Boghossian, P. (1989) 'The Rule-Following Considerations', *Mind* 98: 507-49.(A good summary of recent work on this topic taking up some issues not dealt with here. Contrasts with these discussions in trying to separate the issue of rule-following from the issue of the speaker's knowledge of the rules he is following.)

Budd, M. (1984) *Wittgenstein's Philosophy of Psychology*, London: Routledge. (Chapter 2 provides a clearly written and careful treatment of Wittgenstein's views on meaning and understanding.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell. (Provides a lucid, highly accessible introduction to the problems. Referred to in §§3-4.)

McDowell, J. (1984) 'Wittgenstein on Following a Rule', *Synthèse* 58: 325-63. (Thorough treatment of the issues discussed, including the position outlined in §§3 and 5. Also provides both an interpretation of Wittgenstein's thought and a critical discussion of Kripke and Wright.)

McGinn, C. (1984) *Wittgenstein on Meaning*, Oxford: Oxford University Press.(A clear and concise treatment of rule-following and the private language argument.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(The principal source of the issues discussed here; see especially §§185-289, referred to in §§1-3. Notoriously subtle and elusive writings.)

Wittgenstein, L. (1956) *Remarks on the Foundations of Mathematics*, Oxford: Blackwell. (Section 6 includes key passages on rule-following. Intriguing but not straightforward. Referred to in §2.)

Wright, C. (1980) *Wittgenstein On the Foundation of Mathematics*, London: Duckworth.(See chapters 2 and 11 for Wright's early views on rule-following, mentioned in §4.)

Wright, C. (1989a) 'Wittgenstein's Rule-following Considerations and the Central Project of Theoretical Linguistics', in A. George (ed.) *Reflections on Chomsky*, Oxford: Blackwell.(Difficult though rewarding reading, setting out the account described in §§3 and 6.)

Wright, C. (1989b) 'Critical Notice of *Wittgenstein on Meaning* by Colin McGinn', *Mind* 98: 289-305.(Section 3 contains a penetrating discussion of the principal themes and arguments from the key passages in Wittgenstein's *Philosophical Investigations* (1953) and *Remarks on the Foundations of Mathematics* (1956).)

Meaning and truth

Analytic philosophy has seen a resurgent interest in the possibility of explaining linguistic meaning in terms of truth, which many philosophers have seen as considerably more tractable than meaning. The core suggestion is that the meaning of a declarative sentence may be given by specifying certain conditions under which it is true. Thus the declarative sentence 'Venus is red' is true just in case the condition that Venus is red obtains; and this is exactly what the sentence means.

As it stands, however, this suggestion provides us with no explanation of the meanings of the words and phrases that make up sentences, since in general they are not expressions that have truth-conditions. (There are no conditions under which the word 'Venus' is true.) Furthermore, it needs to be supplemented by some method of circumscribing the truth-conditions that embody the meanings of declarative sentences, since there are many conditions under which any given sentence is true: 'Venus is red' is true not merely when Venus is red, but also, for example, when Venus is red and 7 + 5 = 12; but it does not mean *that Venus is red and 7 + 5 = 12.*

Evidently the first problem can be solved only by finding other semantic properties which indicate the meanings of words and phrases. For example, it is sometimes thought that the meaning of a name can be specified by saying what it refers to; and that of a predicate by saying what it is true of. But notice that since the meaning of a declarative sentence can be grasped by first grasping the meanings of its basic components, meaning-indicating ascriptions of semantic properties to those components must entail a meaning-indicating statement of its truth-conditions. Semantic properties such as 'referring to' and 'being true of' satisfy this requirement, at least in the context of what is sometimes called a 'truth theory' for a language.

This still leaves the problem of how to circumscribe the right meaning-indicating statement of truth-conditions for declarative sentences. Indeed we now have a further problem. For if the meanings of the components of sentences are not stated directly, but merely in terms of what they refer to or are true of (say), then we must also find a way of determining which of the many ways of specifying what they refer to, or the conditions under which they are true of something, is meaning-indicating. These problems may arguably be solved by placing an appropriate truth theory for a language in a setting that allows us to appeal to the general psychology of its speakers.

Attempts to elucidate meaning in terms of truth-conditions induce a plethora of further problems. Many are a matter of detail, concerning the kinds of properties we should associate with particular idioms and constructions or, equivalently, how we are to produce truth theories for them. As a result of Tarski's work, we have a good idea how to do this for a wide range of categories of expressions. But there are many which, superficially at least, seem to resist straightforward incorporation into such a framework. More general difficulties concern whether truth should be central at all in the analysis or elucidation of meaning; two objections are especially prominent, one adverting to antirealist considerations, the other to the redundancy theory of truth.

1 The core suggestion and some problems

Three of the most influential approaches to the understanding of linguistic meaning in contemporary analytic philosophy correspond to three different, though overlapping, roles of language: its role in communication; its role in reasoning; and its role in making statements. According to the first of these approaches, linguistic meaning should be analysed or elucidated in terms of the communication intentions of speakers (see Communication and intention; Meaning and communication). According to the second, it should be characterized in terms of the conceptual or inferential role of expressions of the language (see Semantics, conceptual role). And the third, at least in the form in which it has been most fully articulated, requires meaning to be analysed or elucidated in terms of the truth-conditions of sentences whose characteristic role, at least in context, is to make statements. It is the third approach which is the subject of this entry.

The core suggestion (which we owe to Frege (1964)) is that the meaning of a declarative sentence is given by certain conditions under which it is true. Thus 'Venus is red' is true just in case Venus is red; and that Venus is red is precisely what the sentence means. However, this fact leaves us in the dark about the meanings of expressions which are not sentences. For example, none of the words that makes up the sentence 'Venus is red' is the kind of expression that can have truth-conditions. Furthermore, there are countless other conditions under which the

sentence is true that do not reveal its meaning - that Venus is red and $7 + 5 = 12$, for example. So the core suggestion must be supplemented not only by some way of articulating word meaning, but also by a method of picking out those truth-conditions of declarative sentences that embody their meanings - a method which does not make reference to meaning itself. I shall call the problem of articulating word meaning 'Problem A', and the problem of picking out the right truth-conditions of sentences 'Problem B'.

Let us begin with Problem A. It is clear that to solve it we must find further semantic properties which can plausibly indicate the meanings of non-sentential expressions, just as particular truth-conditions are meant to indicate the meanings of declarative sentences. For example, it is plausible to claim that the meaning of a name can be specified by saying what it refers to, and that of a predicate by saying what it is true of: the meaning of the name 'Venus' may be specified by stating that its referent is Venus; and that of the predicate 'is red' by stating that it is true of something if and only if that thing is red.

However, two difficulties arise with this proposed solution to Problem A. The first - which I shall call D(i) - is an analogue of Problem B. This is that there are many ways of stating the referent of a name, and many ways of stating what predicates are true of, but not all of them are meaning-indicating. Thus, although we may say that the referent of 'Venus' is Venus, we may also say that it is Hesperus, or the thing that is both Venus and either green or not green; but only the first statement of its referent specifies its meaning. Equally, although we may say that 'is red' is true of something if and only if that thing is red, it is also true of something if and only if the thing is red and $2 + 2 = 4$, and so on; but again only the first biconditional specifies its meaning.

The second difficulty, D(ii), arises from the fact that anyone who understands the components of a given sentence is generally in a position to understand the sentence itself; no further information is needed. For example, if I understand both 'Venus' and 'is red', I can immediately go on to understand the sentence 'Venus is red'. Now this is tantamount to saying that the meanings of sentences are determined by the meanings of their components (see Compositionality). Thus the meanings of 'Venus' and 'is red' determine that of 'Venus is red'. But if this is right, then any correct specifications of the meanings of a sentence's components should entail meaning-indicating statements of the truth-conditions of the sentence itself. In particular, the specifications of the meanings of 'Venus' and 'is red' should entail the meaning-indicating statement of the truth-conditions of 'Venus is red', namely that 'Venus is red' is true if and only if Venus is red. (If not, the putative meaning-specifications of 'Venus' and 'is red' must have left something out.) The problem, however, is that even in this simple case it is not immediately clear how to ensure this.

2 Truth theories

It is convenient to tackle difficulty D(i) after we have dealt with D(ii). We can then deal with D(i) along with Problem B (the corresponding difficulty for sentences). Turning to D(ii), then, how are we to ensure that correct meaning-specifications for a sentence's components entail a correct specification for the sentence itself? Here it is customary to appeal to what is often called a 'truth theory' for a language. For it is precisely by employing such a theory that statements of truth-conditions of sentences of a language can be deduced from statements about the semantic properties of their components (see Tarski, A.).

What is a truth theory for a language? Ordinarily, it is taken to consist of three parts: syntactic, semantic and logical. The syntactic part comprises a set of clauses stating how sentences of the language are constructed out of the vocabulary of the language. These clauses allow us to construct a description of any sentence of the language which shows how it is composed out of its vocabulary. (Such descriptions are often called 'structural descriptions'.) For example, in a very simple subject-predicate language - call it L - whose vocabulary consists of the English names 'Venus' and 'Polaris' and the English predicates 'is red' and 'is bright', four sentences are constructible. Letting ' $x^\wedge y$' mean the result of concatenating x with y, we may describe these sentences structurally as follows: 'Venus' $^\wedge$ 'is red', 'Polaris' $^\wedge$ 'is red', 'Venus' $^\wedge$ 'is bright' and 'Polaris' $^\wedge$ 'is bright'. (This notation is rather clumsy, so, for ease of reading, I shall use quotation marks both in the ordinary way and to indicate structural descriptions.)

Next, the semantic part comprises a set of clauses specifying various semantic properties of the vocabulary of the language, together perhaps with a number of semantic properties of complex expressions articulated as a function of the semantic properties of their components. In tandem with the third part, which consists of a number of logical

and set-theoretic principles, these clauses will determine the truth-conditions of sentences of the language. It is this last feature that allows us to see in principle how we might solve D(ii).

Consider the language *L* again. Our aim is to find semantic properties of its names and predicates which will determine the truth-conditions of sentences containing them. And if what we have said so far is right, this should involve a statement of what (according to the theory) the names refer to and the predicates are true of. Thus we might have

(1) The referent of 'Venus' = Venus;
(2) The referent of 'Polaris' = Polaris;
(3) 'is red' is true of something iff that thing is red;
(4) 'is bright' is true of something iff that thing is bright.

(Note that strictly speaking the semantic properties here should be relativized to *L*, for in other languages its names and predicates may refer to or be true of very different things. Suppressing the language parameter, however, makes the axioms more readable. The same is true of all semantic properties considered in this entry.)

At this point, of course, we still do not know how these clauses determine specific truth-conditions for sentences. But we may remedy this by adding the following general clause relating the truth of *L*-sentences to the semantic properties of their components:

(5) For all names γ and predicates *F*, ' γF ' is true if and only if *F* is true of the referent of γ.

For assuming that the third, logical part of the truth theory includes elementary first-order logic, it is then a simple matter to deduce truth-conditions for the sentences. For instance, we may easily deduce that 'Venus is red' is true if and only if Venus is red. For (5) will entail that

'Venus is red' is true if and only if 'is red' is true of the referent of 'Venus';

(1) will entail that

'is red' is true of the referent of 'Venus' if and only if 'is red' is true of Venus;

and (3) will entail that

'is red' is true of Venus if and only if Venus is red.

Similar reasoning will enable us to deduce that 'Polaris is red' is true if and only if Polaris is red, and so on for the other two.

We are thus able to deduce a statement of what are in fact meaning-indicating truth-conditions for each of the sentences of *L* on the basis of axioms which assign semantic properties to their components, along with a further axiom which characterizes their modes of combination. Of course, this only solves our difficulty for the very limited language *L*; but it does give a hint of how to solve it more generally.

3 Providing correct meaning-specifications

Let us turn now to Problem B - how to circumscribe in a non-circular manner those truth-conditions of declarative sentences which embody their meanings - together with D(i), the corresponding problem for sentence components. And let us begin by introducing a common abbreviation: let us call a biconditional of the form '*S* is true iff *p*' a '*T*-sentence'. Since *T*-sentences evidently express the truth-conditions of the declarative sentences mentioned on their left-hand sides, it follows that to deal with Problem B, we need a non-circular way of singling out those *T*-sentences which genuinely indicate the meanings of the declarative sentences from those which do not. One suggestion would be to require that for '*S* is true iff *p*' to be a meaning-indicating *T*-sentence, '*p*' must be a translation of *S*. But although this would certainly ensure that the *T*-sentence correctly indicated the meaning of the relevant sentence, it would not provide a non-circular elucidation of its meaning, since the notion of translation itself directly presupposes the notion of meaning: when *A* and *B* are distinct expressions, *A* is a translation of *B* if and only if *A* and *B* have the same meaning.

A second suggestion is to replace the notion of translation with more general anthropological considerations relating to interpretation. Speech and writing are activities which take place against a backdrop of broadly rational activity, and can be interpreted in the light of such activity with a view to making maximum sense of speakers' behaviour. Consequently, it should be possible in principle to determine whether speakers are saying, for example, that Venus is red when they utter a particular sentence *S* by assessing whether that would be the most sensible thing for them to be saying in the circumstances, and in the light of all their other behaviour. And if this is what they typically or characteristically say when they utter *S*, then (according to the second suggestion) we may identify this with the appropriate truth-condition for *S*.

This prompts the following thought: '*S* is true iff *p*' is correctly meaning-indicating if, other things being equal, the supposition that speakers are saying that *p* in characteristic utterances of *S* makes maximum sense of their behaviour. But even this might be thought to invite a similar criticism to the one which undermined the first suggestion. For does not the use of speech acts such as saying just bring in meaning by the back door? It certainly seems to if saying that *p* is uttering something that means that *p*. But much depends on whether a deeper account of saying - and other speech acts - is possible (see Radical translation and radical interpretation).

A third suggestion would be to forgo such use of speech acts in the characterization of meaning, and focus on speakers' knowledge of truth-conditions. It is clear that those who understand the sentence 'Venus is red' will know not merely that 'Venus is red' is true if and only if Venus is red, but that speakers are expected to know this. Moreover, it also seems clear that they will know that this is the strongest thing speakers are expected to know about the truth-conditions of 'Venus is red'. They will know, for example, that speakers are not expected to know that 'Venus is red' is true if and only if Venus is red and $7 + 5 = 12$. But this suggests the following proposal: that '*S* is true iff *p*' is meaning-indicating if any speaker knows that *S* is true if and only if *p*, that speakers are expected to know this, and that this is the strongest thing speakers are expected to know about the truth-conditions of *S*. Our *T*-sentence which fails to specify the meaning of 'Venus is red' - 'Venus is red' is true if and only if Venus is red and $7 + 5 = 12$ - is then ruled out (see Meaning and understanding).

Turning now to D(i), what of the components of sentences? How can we ensure that their meaning-specifications are correct? A thought that has occurred to several writers is to let whatever is the most satisfactory method of circumscribing the meaning-specifying truth-conditions of sentences do all the work: to say, for example, that the meaning of a non-sentential expression of a language can be specified by its clauses in any true truth theory that meets the above kinds of knowledge-expectation conditions for its sentences. For this to be adequate, however, every true truth theory for a language that meets those conditions would have to assign to those expressions the same or equivalent semantic properties. Unfortunately this not true.

To see this let us return to the truth theory of our language *L*. As we saw in §2, it is possible to use that truth theory to derive a *T*-sentence for each sentence of *L*; and each such *T*-sentence plainly satisfies the relevant knowledge-expectation conditions. What is more, its axioms plausibly specify the meanings of each of *L*'s basic expressions. There are, however, other truth theories for *L* that meet the knowledge-expectation conditions for its sentences, but which do not specify the meanings of its basic expressions. For example, consider a truth theory for *L* which had the following axiom governing 'is red' instead of (3):

(6) 'is red' is true of something if and only if it is red and is either green or not green.

It would certainly be possible to derive all the correct meaning-indicating *T*-sentences for sentences of *L* on the basis of this axiom and the axioms governing the other basic expressions of the sentences. But (6) does not specify the meaning of 'is red' correctly.

It is true that we could rule out such cases by requiring that meaning-indicating *T*-sentences be proved in certain canonical ways - for instance, in the way we derived the *T*-sentence for 'Venus is red' in §2. For a proof that differed from that derivation only in using (6) instead of (3) would result in a *T*-sentence that did not satisfy the knowledge-expectation conditions, namely: 'Venus is red' is true if and only if Venus is red and is either green or not green. But this immediately raises the question why those ways should be thought of as canonical; and since the natural answer is that they ensure that the axioms governing the sentence's basic components specify their meanings correctly, it would seem that we must have some independent grip on what it is for those axioms to be so constrained as to have this property. The question is what provides us with that independent grip.

At this point, we may make a move analogous to the one made earlier in the case of sentences. There we noted that understanding the sentence 'Venus is red' involves knowing that it is true if and only if Venus is red, that speakers are expected to know this, and that this is the strongest thing speakers are expected to know about its truth-conditions. Similarly, we may say that understanding the predicate 'is red' involves knowing it is true of something if and only if that thing is red, that speakers are expected to know this, and that this is the strongest thing speakers are expected to know about the conditions under which it is true of something. (They will know, for example, that speakers will not be expected to know that it is true of something if and only if that thing is red and is either green or not green.) Equally, understanding the name 'Venus' will consist in knowing that it refers to Venus, that speakers are expected to know this, and that this is the strongest thing speakers are expected to know about what it refers to. Notice again that this solves our difficulty only for the very limited language L; but we may expect it to be soluble quite generally by means of similar kinds of knowledge-expectation conditions relating to whatever clauses govern other expressions and their modes of combination.

It should be noted that these kinds of knowledge-expectation conditions appear to put a considerable burden on the ordinary speaker: can we really picture them as part of the stock of knowledge possessed by speakers of a natural language such as English? At this point, if we are not to jettison the present approach, there seem to be three strategies. One is to claim that the truth-theoretic clauses are correct, to deny that we have knowledge of them (in the ordinary sense), but to insist that we bear some kind of analogous cognitive relation to them (yet to be fully explicated). The second is again to claim that they are correct, but to try to justify the ascription of tacit knowledge of them in some way. And the third is to see the truth-theoretic clauses as stepping stones to more convincing clauses which we can uncontroversially accept that we know.

4 Recalcitrant idioms and general difficulties

We turn next to a number of problems for the truth-conditional approach to meaning. In the light of Tarski's work, it is now clear how to accommodate within a truth-theoretic framework declarative sentences made up of names, predicates, truth-functors and standard quantifiers. But natural language seems much richer than this. We here consider two idioms which cause problems for the framework, and which in their different ways may require it to be modified. Then we consider two more general difficulties. It should be noted that there are various further issues - such as how the truth-conditional approach melds with doctrines of logical form, transformational grammar or validity - which, for reasons of space, we cannot discuss here.

Until now we have simplified matters by restricting ourselves to languages whose only sentences are declarative. This restriction is convenient, since - ignoring context-sensitivity - it means that all the sentences in the relevant languages have truth-conditions; but it is not warranted by natural languages, which contain sentences of many different moods, including imperatives, interrogatives and optatives. Such sentences do not appear to be truth-evaluable - 'Open the door, Smith!' or 'Where's the butter?' do not look like the kinds of sentence that can be true or false. Yet if they lack truth-conditions, it seems puzzling how they can be accommodated within a truth-theoretic framework.

One response is to add a further component to the semantic theory - what is sometimes called a theory of 'force' (see Pragmatics §8). This states what kind of speech act a given sentence is conventionally or typically geared to express; and it transforms the sentence into a corresponding declarative sentence which can then serve as input to the truth theory. For example, the theory of force for English would tell us that 'Open the door, Smith!' is a command, and that its corresponding declarative sentence is 'Smith opens the door'.

A second response is to introduce a number of further semantic values analogous to truth for each of the new types of sentence: compliance, for instance, for imperatives and optatives. Such values could be integrated into appropriate truth theories by means of such axioms as: $S!$ is complied with iff S is true, where $S!$ is the imperative form of a declarative sentence S. And to ensure that we obtain meaning-indicating compliance conditions, we could again use knowledge-expectation conditions, similar to those proposed in the case of declarative sentences, but with compliance replacing truth. Note, however, that both responses have great difficulty accommodating a sentence such as 'Open the door, Smith, and I'll put the kettle on' which contains more than one mood. Associated with this sentence, there seems no single speech act, nor any straightforward value analogous to truth.

Next we come to sentences whose truth-conditions can vary from one context to another. Such sentences, which

typically contain context-sensitive expressions such as personal pronouns, demonstratives, temporal and spatial indexicals, tense indicators and so on, evidently give rise to a problem for the programme of accounting for meaning in terms of truth-conditions. For although the truth-conditions of such sentences can vary from context to context, their meaning need not. The sentence 'I helped that man yesterday' means the same thing whoever utters it, whenever it is uttered and whoever the man is; but its truth-conditions will vary with the utterer, the time of utterance and the man indicated. Evidently, this is because in any ordinary way of taking the term 'meaning', the meanings of expressions such as 'I', 'yesterday' and so on are given by such general rules as that 'I' refers to the speaker, 'yesterday' refers to the day before the day of utterance and so on; but such meanings in themselves do not exhaust the contribution the expressions make to the truth-conditions of sentences containing them.

An amendment to the account is clearly needed. In particular, although we may take the truth-conditions of sentences to be identifiable with their meaning when they have context-free truth-conditions, we must identify those of context-sensitive sentences with a notion of content much stronger than ordinary meaning - roughly, with what the sentences express when the context has been fixed. This means that the semantic theory will in effect have to be divided into two parts: what supplies the context-free information relevant to truth-conditions; and what states the information about specific contexts that determines particular truth-conditions once the relevant context-free information has been supplied. The second part falls into the area of pragmatics (see Pragmatics; Demonstratives and indexicals).

Is this adequate? We here restrict ourselves to noting one general difficulty with the proposal. It is likely that most context-free semantic axioms will involve relativized semantic predicates. The pronoun 'he', for instance, might be governed by an axiom such as:

(7) For all p, q, the referent of 'he' as uttered by p to indicate $q = q$.

However, it is hard to obtain content-specifying instances of this which satisfy the appropriate knowledge-expectation conditions involving reference (see §3). For suppose I am the speaker and the person I am indicating is Quine. Then to display the content of 'he' in my mouth, we need a metalinguistic expression synonymous with it - 'a', say. This would give us:

The referent of 'he' as uttered by me to indicate Quine = a.

But even if we had such an expression, we would still face the problem that not all English speakers, even those in my immediate linguistic circle, need know of the referent of 'he' in my mouth that it is a. Indeed, I may be the only person who knows this. In such extreme circumstances, we could restrict the class of speakers to me. But it would then have to be the case that the strongest thing I expect myself to know about its referent is that it is a. And unfortunately there is no reason to think that this is so.

Finally we turn to two general problems for the programme of elucidating meaning in terms of truth-conditions, both of which have considerable contemporary significance. The first, antirealist difficulty focuses on sentences whose truth-conditions we cannot decide. According to one model of understanding, we can acquire and manifest an understanding of a sentence only if we can recognize whatever meaning-indicating condition is associated with it as obtaining when it does obtain and as not obtaining when it does not. But for sentences whose truth-conditions we cannot decide, the supposition that truth-conditions are meaning-indicating conflicts with this requirement. So in the absence of another model, the attempt to elucidate meaning in terms of truth-conditions must be jettisoned (see Intuitionistic logic and antirealism §6; Realism and antirealism §4).

The second difficulty concerns the definition of truth. It is clear that an extant notion of truth is needed to characterize meaning in the way we have been envisaging: we cannot use truth theories like the one described in §2 as analyses of truth, if only because they leave it unsaid how to apply the notion of truth to new languages. However, one of the most plausible independent accounts of truth, the 'redundancy theory' (see Truth, deflationary theories of), is arguably incompatible with any attempt to characterize meaning in terms of truth. For on one natural construal, this theory requires that understanding what it is for 'Venus is red' (say) to be true is tantamount to understanding what it is to say that Venus is red by using the sentence 'Venus is red'. But to understand this, we already need to know the meaning of 'Venus is red'.

Both objections seem to require that we replace truth by some other notion in an elucidation of meaning. One that

appeals to antirealists is that of warranted assertibility, since the conditions under which one would be warranted in asserting a sentence are arguably recognisable as obtaining when they obtain and as not obtaining when they do not. Other options would be to elucidate meaning in terms of communication intentions or conceptual role.

See also: Davidson, D.; Logical and mathematical terms, glossary of

STEPHEN G. WILLIAMS

References and further reading

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press.(These essays present the most influential contemporary defence of a truth-conditional account of meaning; see in particular Essays 2, 9 and 10. Essay 8 discusses the problem of non-declarative sentences; his own account is different from those noted in §4.)

Davies, M.K. (1981) *Meaning, Quantification, Necessity*, London: Routledge. (Includes truth theories for a range of English idioms; chapter 4 has a helpful discussion of tacit knowledge.)

Dummett, M.A.E. (1978) *Truth and Other Enigmas*, London: Duckworth.(Includes a number of seminal essays on antirealism; see in particular Essays 1, 12 and 21. Essay 1 and the Preface argue for the incompatibility of the redundancy theory of truth and the truth-conditions theory of meaning.)

Evans, G. and McDowell, J. (eds) (1976) *Truth and Meaning*, Oxford: Oxford University Press.(Includes an excellent introduction by the editors.)

Frege, G. (1964) *The Basic Laws of Arithmetic*, trans. and ed. and with an intro. by M. Furth, Berkeley, CA: University of California Press.(This is a translation of parts of Frege's *Grundgesetze der Arithmetik*. The suggestion that the meaning of a declarative sentence may be given by the conditions under which it is true is to be found, wrapped up in Frege's own terminology, in §32.)

Higginbotham, J. (1991) 'Truth and Understanding', repr. in C. Peacocke (ed.) *Understanding and Sense*, Aldershot: Dartmouth, 1993, vol. 1.(Presents detailed elucidation of some of the material in §3; in particular suggests circumscribing truth-conditions and so on in terms of knowledge-expectation conditions.)

Larson, R. and Segal, G. (1995) *Knowledge of Meaning: An Introduction to Semantic Theory*, Cambridge, MA: MIT Press.(An excellent survey of truth theories for many English idioms; it also indicates how to integrate this material with current thinking in linguistics. The first three chapters include accounts of some of the basic ideas underlying the truth-conditional approach to meaning; chapter 13 discusses tacit knowledge.)

Lepore, E. (ed.) (1986) *Truth and Interpretation*, Oxford: Blackwell.(This volume includes important essays on and within the framework of truth-conditional theories of meaning. Particularly helpful are those by Higginbotham, Davidson and Dummett.)

McDowell, J. (1977) 'On the Sense and Reference of Proper Name', repr. in M. Platts (ed.) *Reference, Truth and Reality*, London: Routledge, 1980; and in C. Peacocke (ed.) *Understanding and Sense*, Aldershot: Dartmouth, 1993, vol. 1.(Suggests coupling a truth-conditional account with a theory of force in the way indicated briefly in §4. Also fills out material in §3.)

Peacocke, C. (ed.) (1993) *Understanding and Sense*, 2 vols, Aldershot: Dartmouth. (These volumes also include important essays on and within the framework of truth-conditional theories of meaning. In addition to those by Higginbotham and McDowell already recorded, the essays by Davidson on truth, by McDowell on antirealism and by Quine and Davies on tacit knowledge are notable.)

Platts, M. (ed.) (1980) *Reference, Truth and Reality*, London: Routledge.(Includes useful essays on Tarski's account of truth, by Field and McDowell.)

Wright, C. (1987) *Realism, Meaning and Truth*, Oxford: Blackwell, 2nd edn, 1993. (Contains important essays on the antirealist critique of truth-conditional semantics. Introduction is particularly helpful.)

Yourgrau, P. (ed.) (1990) *Demonstratives*, Oxford: Oxford University Press.(A useful selection of essays on context-sensitivity.)

Meaning and understanding

The existence of a close connection between the notions of meaning and understanding can hardly be denied. I may be said to understand you, on a given occasion of utterance, when I know what you then meant - or, at least, when I know the meaning of the words that you then uttered. An important and influential school of thought within the philosophy of language goes much further than these platitudes, however. Its members adhere to the view that questions about meaning are essentially questions about understanding: 'a model of meaning is a model of understanding'. Their approach contrasts with that of those who expect an account of meaning to elucidate the nature of understanding only indirectly - perhaps by explaining meaning in terms of truth, inference, synonymy or self-expression, and only then explaining understanding as the correct recovery of meaning.

1 Dummett's account

The text of this article sketches and constrasts the views of the school's two leading members: Dummett and Davidson. The dictum 'a model of meaning is a model of understanding' was formulated by Michael Dummett. He proceeds to gloss 'a model of understanding' as 'a representation of what it is that is known when an individual knows the meaning' of a possible object of understanding - as it might be, a word, a construction, or a whole language (1973: 217). To provide such a model is to spell out the kind of knowledge that constitutes, or at least sustains, various kinds of linguistic understanding. Dummett's view is that the central philosophical debates about the nature of meaning are properly understood to be debates about the character of such knowledge. Slogans such as 'The meaning of a sentence is the method of its verification' or 'The meaning of a sentence consists in its truth-conditions' mean (if they mean anything at all) that to *understand* a sentence is to know how it may be verified or to know under what conditions it is true. In emphasizing that the central philosophical questions about meaning are questions about understanding, Dummett was self-consciously departing from a prior tradition (of which the most celebrated manifestation is Quine's 'Two Dogmas of Empiricism'); this tradition took the central problem in the area to be that of spelling out the conditions for two expressions to mean the same thing (see esp. Dummett 1981: 74).

Although Dummett canvasses various accounts of knowledge of meaning, it is a presupposition of his discussion that the knowledge in question is propositional. To attempt to produce a model of understanding is to attempt to produce a *theory*, knowledge of which constitutes or sustains possession of a linguistic capacity. In saying so much, Dummett is not committed to denying that an understanding of a word or sentence is in the first instance a matter of knowing how to use it. However, he is committed to denying that this formulation provides a complete account of understanding, he argues that we need a more detailed specification of what such know-how consists in (1973: 222-3). Dummett is also clear that the propositional knowledge in question will in general be implicit. That is to say, it is knowledge that a subject may possess even though he is unable to express the proposition thereby known. From this observation, Dummett derives a consequence that crucially constrains the sort of knowledge which he thinks might constitute or sustain understanding. 'Implicit knowledge', he says, 'cannot meaningfully be ascribed to somebody unless it is possible to say in what the manifestation of that knowledge consists: there must be an observable difference between the behaviour and capacities of someone who is said to have that knowledge and someone who is said to lack it' (1973: 217). In particular, then, the propositional knowledge, implicit possession of which is said to underlie understanding, must be manifestable in behaviour.

The significance of this requirement emerges when we consider the two models of understanding that Dummett centrally considers. Both of these models take to be central the problem of characterizing our understanding of declarative sentences, and according to the first of them - which Dummett labels a 'realist' or 'truth-conditional' model - such understanding 'consists in our grasp of [the sentence's] truth-conditions, which determinately either obtain or fail to obtain, but which cannot be recognized by us in all cases as obtaining whenever they do' (1972: 23). Thus, our understanding of the sentence 'A city will never be built at the North Pole' is taken by the realist to consist in our knowing that it is true just on condition that a city will never be built at the North Pole, a condition that may obtain without our ever knowing so (Dummett 1959: 16-7). This, Dummett argues, is rendered problematical by the requirement of manifestability. 'It is quite obscure', he writes, 'in what the knowledge of the condition under which a sentence is true can consist, when that condition is not one which is always capable of being recognized as obtaining' (1973: 224). The crucial claim, in other words, is that there is

something opaque about the attribution 'Tom knows that sentence *S* is true just on condition that a city will never be built at the North Pole'. Because the occasion may never arise on which somebody is granted the opportunity to give conclusive reasons for accepting the sentence as true or rejecting it as false, 'the knowledge which is being ascribed to one who is said to understand the sentence is knowledge which transcends the capacity to manifest that knowledge by the way the sentence is used'. Accordingly, the truth-conditional model of understanding cannot be one 'in which meaning is fully determined by use' (1973: 225).

Whether this really is a consequence of the truth-conditional model - and whether, if it is, it is sufficient to show that the realist model must be rejected - are questions that remain at the forefront of contemporary debate in the philosophy of language. As he made clear in 'Realism: A Valedictory Lecture' (1992), Dummett is not committed to particular answers. He has, however, articulated an alternative model of understanding which, whatever its other difficulties, is supposed to meet the requirement of manifestability. On this alternative 'antirealist' model, our understanding of a sentence will consist, not in knowing under what conditions it is true, but 'in knowing what recognizable circumstances determine it as true or as false' (1972: 23). Or, to deploy a formulation that Dummett treats as equivalent, it consists in knowing under what circumstances a sentence may be asserted. In addition to meeting the demand of manifestability, this model is supposed to have the merit of leading to a more plausible account of what learning a language involves:

> Thus we learn to assert '*P* and *Q*' when we can assert *P* and can assert *Q*; to assert '*P* or *Q*' when we can assert *P* or can assert *Q*.... We no longer explain the sense of a statement by stipulating its truth value in terms of the truth value of its constituents, but by stipulating when it may be asserted in terms of the conditions under which its constituents may be asserted.
>
> (1959: 17-18)

2 Davidson's account

It may seem as though Donald Davidson is a clear example of a Dummettian realist - someone who takes our understanding of a declarative sentence to consist in knowing the conditions under which it is true. However, while Davidson certainly regards the notion of a truth-condition as central to an account of a subject's linguistic competence, there are nuances in his treatment of these matters that make such a characterization misleading. He begins a celebrated article thus:

> Kurt utters the words 'Es regnet' and under the right conditions we know that he has said that it is raining. Having identified his utterance as intentional and linguistic, we are able to go on to interpret his words: we can say what his words, on that occasion, meant. What could we know that would enable us to do this? How could we come to know it? The first of these questions is not the same as the question what we *do* know that enables us to interpret the words of others. For there may easily be something we could know and don't, knowledge of which would suffice for interpretation, while on the other hand it is not altogether obvious that there is anything we actually know which plays an essential role in interpretation.
>
> (1973: 125)

Davidson is surely right to draw attention to the difference between his central question about knowledge of meaning ('What could we know that would enable us to know what these words mean?') and Dummett's ('What *do* we know when we know what words mean?'). But it is disappointing that he does not give his grounds for doubting that 'there is anything we actually know which plays an essential role in interpretation'. For if there is such actual knowledge, then questions about its character must surely be deemed more urgent and important than the hypothetical question that Davidson chooses to address. We appear, in other words, to need a reason for taking as central the problem upon which he focuses.

Be that as it may, Davidson's own question is a clear one, and his writings contain two rather different answers to it. In his early papers (roughly speaking, those published before 1970, including the influential 'Truth and Meaning' of 1967), we are told that what would suffice for interpreting a speaker *A* at time *t* is knowledge of a truth theory that is interpretative for *A* at *t*. A truth theory, as Davidson conceives it, comprises statements (*T*-theorems) to the effect that a sentence is true (as potentially uttered by such-and-such a speaker at this or that time) just in case a certain condition is met. And such a theory will be interpretative, for a speaker at a time, if (intuitively) the condition that the theory associates with each sentence gives the meaning that it then has in that

speaker's mouth. Plainly, this answer is of little interest without some independent specification of the condition for a theory to be interpretative (independent of the notion of meaning), which Davidson hopes to provide via the figure of the radical interpreter. Such an interpreter is an investigator who constructs a truth theory for a speaker (at a particular time) in accordance with certain general interpretative maxims or constraints, chief among which is the principle of charity (see Davidson 1973; Lewis 1974; Charity, principle of §4; Radical translation and radical interpretation §7-10). The idea, then, is that it is such maxims as these that give content to the requirement that a theory be interpretative, and thereby give content to the notion of meaning.

The model of understanding and meaning that results involves a substantial holism. The condition for a sentence S to mean that p as a speaker A uses it at time t, is that a truth theory that is interpretative for A at t (that is, a truth theory that optimally meets the constraints under which the radical interpreter operates) should contain among its T-theorems the result 'S is true (as potentially uttered by A at t) just in case p'. This makes it clear that the condition for a sentence to have a certain meaning involves reference, not just to the components of S, but to the totality of sentences that the relevant speaker is then disposed to utter (see Davidson 1976: 175). This holism does not alarm Davidson, but he has confessed to an important error in his earlier answer to his question. 'My mistake', he wrote in 1976, 'was to overlook the fact that someone might know a sufficiently unique theory [that is a theory that optimally satisfies such constraints as the principle of charity] without knowing that it was sufficiently unique' (1976: 173). He duly amends his account of what suffices for interpretation to correct this mistake: 'what somebody needs to know [in order to interpret speakers of a language L] is that some [sufficiently unique] T-theory for L states that… (and here the dots are to be replaced by a T-theory)' (1976: 174).

The occurrence in this formulation of the intentional and linguistic verb 'to state' may lead one to wonder whether the account is invoking a notion to which it is not entitled (compare Forster 1976). After all, if we are allowed to use a verb like 'state' in specifying the knowledge that would suffice for understanding, then why should we not bypass the complexities of truth theories completely, and say simply that what suffices for interpreting a speaker is knowing what he would be stating by affirming declarative sentences, what he would be asking by uttering interrogative ones, and what he would be ordering by uttering imperative sentences? Davidson, however, has a convincing answer to this question. Although the use of the verb 'state' (or, as he eventually prefers, 'entail') 'brings in an appeal to a specifically linguistic notion', it is a notion that has been 'elicit[ed] by placing conditions on a theory of truth' (1976: 178). That is to say, its content must be understood in relation to the conditions for a truth theory to be interpretative.

All the same, one might wonder whether a more direct method of eliciting (or explaining) such a notion was not available. Instead of considering truth theories - and regarding the principle of charity and its ilk as providing conditions for a truth theory to be interpretative - why should we not consider theories entailing such claims as that:

By affirming S at t, A would state that p

and regard the relevant principles as conditions for such a theory to be *true*? Davidson's answer appears to be that certain logical difficulties arise in the latter case that are absent from the former. In an early paper, Davidson wrote that 'when we can regard the meaning of each sentence as a function of a finite number of features of the sentence, we have an insight not only into what there is to be learned; we also understand how an infinite aptitude can be encompassed by finite accomplishments' (1965: 8). The context makes it clear that the present occurrence of 'when' means 'when and only when' and the suggestion is that such insight will be gained only when we have an axiomatization - if not finite, then at least recursive - of the body of knowledge in question. Now Davidson also suggests that a theory whose theorems are in the form:

By affirming S at t, A would state that p

will resist axiomatization. 'It is reasonable to expect', he writes, 'that in wrestling with the logic of the apparently non-extensional "means that" [and, we might add, the equally non-extensional "would state that"] we will encounter problems as hard as, or perhaps identical with, the problems our theory is out to solve' (1967: 22).

But are such problems insuperable? Surely one can account for somebody's knowing that Kurt uses 'Es regnet und es schneit' to state that it is raining and snowing by reference to his knowing:

(1) that Kurt uses 'Es regnet' to state that it is raining

(2) that Kurt uses 'Es schneit' to state that it is snowing

and

(3) that $(p)(q)(S)(S')$(If Kurt uses S to state that p, and uses S' to state that q, then he uses 'S und S'' to state that p and q)?

One source of concern about this last item of knowledge may be that the initial pair of quantifiers must be understood substitutionally (see Quantifiers, substitutional and objectual), but it is noteworthy that on Davidson's preferred analysis of such constructions as 'states that', there is a far more acute problem. On his 'paratactic' analysis, what follows the word 'that' in an attribution such as 'Galileo stated that the earth moves' belongs in a semantically separate sentence from the properly attributive claim 'Galileo said that' (see Davidson, D. §6; Propositional attitude statements §3). Applied to (c), then, the substitutional quantifiers '(p)' and '(q)' would reach into a different sentence. Here, the 'direct' approach to theories of meaning seems to take us into uncharted logical waters. Unless Davidson's analysis of indirect discourse can be shown to be wrong, or unless some way can be discovered of charting these waters, we have a reason for answering his question in the way that he recommends.

See also: Meaning and truth; Meaning and verification

IAN RUMFITT

References and further reading

Davidson, D. (1965) 'Theories of Meaning and Learnable Languages', in Y. Bar-Hillel (ed.) *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*, Amsterdam: North Holland; repr. in D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1984, 3-15. (Gives the rationale for requiring an interpretative truth theory to be finitely axiomatizable.)

Davidson, D. (1967) 'Truth and Meaning', *Synthèse* 17: 304-23; repr. in D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1984, 17-36.(The classic statement of Davidson's earlier conception of a theory of meaning.)

Davidson, D. (1973) 'Radical Interpretation', *Dialectica* 27: 313-28; repr. in D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1984, 125-39.(Elaborates Davidson's later conception of theories of meaning.)

Davidson, D. (1976) 'Reply to Foster', in G. Evans and J. McDowell (eds) *Truth and Meaning: Essays in Semantics*, Oxford: Oxford University Press, 33-41; repr. in D. Davidson, *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1984, 171-9.(Davidson compares his earlier and later views, explaining the reasons for the change.)

Dummett, M.A.E. (1959) 'Truth', *Proceedings of the Aristotelian Society* 59: 141-62; repr. in M.A.E. Dummett, *Truth and Other Enigmas*, London: Duckworth, 1978, 1-19. (An early and influential presentation of the idea that to understand a sentence is to know under what conditions it could be asserted.)

Dummett, M.A.E. (1972) 'Postscript to Truth', in M.A.E. Dummett, *Truth and Other Enigmas*, London: Duckworth, 1978, 19-24.(Clarifies the thesis and argument of the previous article.)

Dummett, M.A.E. (1973) 'The Philosophical Basis of Intuitionistic Logic', in H.E. Rose and J.C. Shepherdson (eds) *Logic Colloquium '73*, Amsterdam: North Holland; repr. in M.A.E. Dummett, *Truth and Other Enigmas*, London: Duckworth, 1978, 215-47.(Contains Dummett's clearest and most influential formulation of the requirement that linguistic knowledge should be manifestable in use.)

Dummett, M.A.E. (1981) *The Interpretation of Frege's Philosophy*, London: Duckworth.(The chapter entitled 'Synonymy' compares Dummett's way of formulating the questions that a theory of meaning ought to answer with previous conceptions, notably Quine's.)

Dummett, M.A.E. (1992) 'Realism: A Valedictory Lecture', Oxford: Oxford University Press. (Dummett's discussion of the state of the debate between truth-conditional theories of meaning and their rivals.)

Forster, J.A. (1976) 'Meaning and Truth Theory', in G. Evans and J. McDowell (eds) *Truth and Meaning: Essays in Semantics*, Oxford: Oxford University Press, 1-32.(An influential attack on Davidson's earlier and later theories.)

Lewis, D.K. (1974) 'Radical Interpretation', *Synthèse* 27: 331-44.(Spells out explicitly the conditions under which the radical interpreter operates.)

Quine, W.V. (1951) 'Two Dogmas of Empiricism', *Philosophical Review* 60: 20-43.(Explains the problems

attending the explication of 'synonymous'.)

Strawson, P.F. (1970) 'Meaning and Truth: an Inaugural Lecture', Oxford: Oxford University Press; repr. in P.F. Strawson, *Logico-Linguistic Papers*, London: Methuen, 1971, 170-89.(An influential attack upon the whole school of thought discussed here. Strawson argues that the enterprise of spelling out truth-conditions must be pursued within a theory that explains meaning as relating, in the first instance, to communicative intentions.)

Meaning and verification

The verifiability theory of meaning says that meaning is evidence. It is anticipated in, for example, Hume's empiricist doctrine of impressions and ideas, but it emerges into full notoriety in twentieth-century logical positivism. The positivists used the theory in a critique of metaphysics to show that the problems of philosophy, such as the problem of the external world and the problem of other minds, are not real problems at all but only pseudoproblems. Their publicists used the doctrine to argue that religion, ethics and fiction are meaningless, which is how verificationism became notorious among the general public.

Seminal criticism of verification from around 1950 argues that no division between sense and nonsense coincides tidily with a division between science and metaphysics, as the positivists had claimed. Quine later developed verificationism into a sort of semantic holism in which metaphysics is continuous with science. In contrast, Dummett argues from a reading of Wittgenstein's claim that meaning is use to a rejection of any sort of truth surpassing the possibility of knowledge, and thence to a defence of intuitionistic logic. But the claim that all truths can be known yields in an otherwise innocuous setting the preposterous consequence that all truths actually are known. There are ways to tinker with the setting so as to avoid this consequence, but it is best to conclude by reductio *that some truths cannot be known and that verificationism is false. That in turn seems to show that the prospects for an empiricist theory of meaning are dim, which might well shake a complacent confidence in meaning.*

1 Hume

Perception is the only channel to us from the world whose influx is generally accepted to justify our beliefs about the world. Beliefs figure as units of justification in this empiricist-sounding doctrine. Since justifying something is giving a reason for thinking it true, the units of justification should be the right size for being true or false. Word-sized bits, be they words themselves or their meanings or people's ideas of things, are too small, while paragraph-sized chunks are too big, but sentence-sized stretches are of just the right size and unity for having truth-values. It is natural to take the things believed and justified by experience to be sentence-sized or propositional.

But the unit of Hume's seminal empiricism seems smaller. He begins the *Treatise of Human Nature* (1739-40) with the doctrine that all ideas are derived from impressions. To the extent that Hume made a self-conscious distinction between words and sentences (on the language side), or between concepts and propositions (on the Platonic side), or between ideas and thoughts (on the psychological side), the ideas Hume had in mind seem smaller than sentence-sized. In that way, Hume's empiricism seems narrower than the doctrine with which we began.

Hume claimed ideas are derived from impressions by copying, the copies differing from their originals only in vivacity. Ideas and impressions were thus to be images, the ideas merely fainter reproductions of the impressions. Objections to taking abstract ideas as images predate Hume, but the doctrine at least places ideas in a familiar psychological category. Items in this category are accessible to inner sense or introspection if anything is, and thus are knowable by processes such as those of the five outer senses already empirically certified. Admittedly, images seem too particular to reflect the abstract nature of abstract ideas. But at least acquaintance with images could explain action. If we try to rethink ideas not as images but as capacities or dispositions, we seem left with bare possibilities or potentialities, ungrounded in actualities which could explain action. Faced with a choice between bad theory and none, we often seem to prefer the bad to none; at least bad theory may provide a (risky) basis for getting on, but none provides none. Thinking of ideas on the model of images does not go away.

The doctrine that ideas are images derived from sense impressions by copying can be put to work in two different ways. One might take it for granted that we do indeed have certain ideas we seem to have, and then cast about for impressions from which such ideas might be derived; such a cast might well hook surprisingly unfamiliar epistemic fish. Alternatively, one might start from a catalogue of familiar impressions, and then winnow out ideas that cannot be derived by copying any such impressions; such a purge might well leave a rather sceptical residue.

Hume thought of himself as a sceptic. Causation and the self are two examples of this sceptical pattern in Hume. In

the case of causation, he never seems to have said exactly which impressions are wanting. Kant writes that experience teaches us that a thing is so-and-so, but not that it cannot be otherwise (*Critique of Pure Reason*, 1787: B3). In Hume's terms, we have impressions only of actuality, never of necessity; we see red, but not having-to-be red. But then there is no impression of necessity from which to derive an idea of it, so there is no making sense of causation as necessitation. At most we experience regular succession (with temporal priority and spatial contiguity), so that is all there can be to our idea of causation. Thus Hume's scepticism about necessary connection in nature is linked to his doctrine of ideas. In the case of the self, Hume asserts quite explicitly that there simply is no such thing as an experience of oneself over and above the ideas and impressions revealed to introspection. So, he infers, there can be nothing to the self and our idea of it but the bundle of the contents of consciousness. This bundle theory of the self has always seemed one of the more shockingly sceptical parts of the *Treatise*.

2 Logical positivism

We have now seen how Hume can use his doctrine of impressions and ideas as a sceptical weapon. The Vienna Circle between the two world wars of the twentieth century claimed an intellectual lineage back to Hume, and it is the Circle to whom we credit (or on whom we blame) the verifiability theory of meaning (see Vienna Circle §3). It would be an anachronism to attribute to Hume our preoccupations with language, meaning and reference. But imagine a philosopher already so preoccupied reconsidering Hume's doctrine of impressions and ideas. Many find it difficult to become accustomed to a self-conscious distinction between the meaning of an expression and what a person means by that expression. Without such a distinction at one's fingertips, it is all too easy to slide between the meaning of an expression and the ideas expressed by people using that expression. If our philosopher also admires Hume's empiricism, they may incline towards the doctrine that these ideas must be derived from impressions. The core of empiricism is the view that sense experience is the basic sort of evidence we have for (that is, for verifying) our beliefs about the world. It is then not unnatural for this suspension - of meaning, ideas, impressions and verification - to precipitate out a verifiability theory of meaning, or at least the slogan that meaning is evidence.

The verifiability theory of meaning may not be present in the greatest works of Vienna positivism, such as Carnap's *Aufbau* or his *Logical Syntax of Language* (see Carnap, R. §§1-3). But it does seem to animate his *Scheinprobleme in der Philosophie* (*Pseudoproblems in Philosophy*, 1967) and it is obvious in Ayer's *Language, Truth and Logic* (1936), an enormously noticed popularization of logical positivism (see Ayer, A.J. §§2-3). In so far as the positivists had verificationism, they used it as a weapon of criticism. Hume's scepticism can be taken as criticism of complacent confidence in causes and the self, but it would be a distortion to assimilate verificationist criticism without further ado to Hume's scepticism. On the contrary, Carnap's *Pseudoproblems* shows him using verificationism to criticize certain sceptical lines of thought, the problem of the external world and the problem of other minds.

Descartes (1641) denied the pretensions of sense experience to be utterly basic to knowledge and its justification. To that end, he asks how one knows one is not asleep and dreaming right now. Come to that, perhaps one has been asleep and dreaming all of one's life, and all one's experience, instead of reporting the world of matter it seems to present, is instead merely an interminable hallucination spun by oneself out of whole cloth. This difficulty is the problem of the external world (see Descartes, R. §9). Indeed, perhaps there is no matter at all, but only one's mind and the endless illusions one foists on oneself. Here we have two incompatible hypotheses. One says that there is the matter out there independent of us that there seems to be, subject to the laws of nature. The other says that there is no such matter, but only the hallucination of it one foists on oneself. It is crucial to Descartes' case that the sense experience forecast by these two hypotheses be indistinguishable. For we are to be sure that the matter hypothesis is true and the dreaming hypothesis is false, and since sense experience cannot separate the two, there must be some other way of knowing, perhaps even more basic than experience, since it is what entitles us to take our experience to support the matter hypothesis. Prominent along that way are the *Cogito* and the ontological proof of the existence of God. Many after Descartes are not persuaded that his way leads anywhere. So they are left without an answer to how they know they are not asleep and dreaming.

In the way one feels one's own headache and observes its location, extent and intensity, there is nothing remotely like feeling someone else's pain or apprehending another's experience. Still less can one see or hear or smell or touch or taste another's experience and come to know it in that way as they do. But then for an empiricist there

arises the question of how one knows anyone else even *has* experience and feeling. Why could all one's own experience not be just as it has been and yet all the other human bodies be mindless lumps (acting as if someone with thoughts and feelings and experiences were in there but really) driven only by blind chemical reactions? This is the problem of other minds, and does not seem to go back earlier than John Stuart Mill in the nineteenth century. Mill offered an argument from analogy to solve this problem. One knows first hand from one's own case that it is one's headache that makes one wince just so, and seeing a similar wince on another's face, one infers by analogy a similar cause, namely a headache, in that other. This argument from analogy proceeded from what must be a single case, namely, oneself alone, so can it justify much certainty? Beyond that, the argument seems not even to address the difficulty the problem of other minds may raise for empiricism, namely, how one could know things like another's pain which one simply could not experience.

3 Logical positivism (cont.)

One can make out connections between Hume's doctrine of ideas and the verifiability theory of meaning, but while Hume used his doctrine as a sceptical axe, the positivists used their story of meaning to undercut scepticism. The problems of the external world and of other minds came to figure centrally on the agenda of modern classical philosophy. Philosophical problems rarely if ever receive solutions consensually certified, so it is understandable (even if not quite right) that such problems apparently sit there forever unsolved. In that case, these problems can easily come to look like demonstrations that it is not known whether there is any matter out there and that it is not known whether there is anyone besides oneself. In that way the agenda of modern philosophy can seem to be a compendium of sceptical theses. But the verifiability theory of meaning was used to argue that these so-called problems are not problems deserving solutions, but pseudoproblems deserving only dissolution. If one asks of a bachelor whether his wife is English, both a yes and a no seem bad answers since both seem to confirm a mistake concealed in the question. This trite example is a very simple illustration of how a misconception can present itself as an insoluble problem. Some of what is or was modernist in philosophy was an idea of criticism not as demonstrating the opposite but rather as unmasking, and thus eliminating, fraudulent questions and problems.

How was this to work? Remember that Descartes designed the dreaming hypothesis with the intention that it and the matter hypothesis issue in exactly the same expectations for sense experience; everything would seem the same to one's senses no matter which of the two were true. So by design no sensory experience could settle the question of which is true; both hypotheses are equally well supported by all perceptual evidence. But on the verifiability theory, meaning is evidence. As the old slogan had it, a difference that makes no difference is no difference: claims that cannot be distinguished on the basis of evidence must coincide. So since no sensory experience could discriminate between the matter and the dreaming hypotheses, and since empiricism is right, and sensory experience is the only evidence against which to judge claims about the world, then the two hypotheses have exactly the same evidence, and so, by verificationism, exactly the same meaning. Since they are synonymous, they are not incompatible at all, but as compatible as possible. Hence, there is no question at all of choosing between them, and so, since that question is what the problem of the external world was, there is no such problem. The problem has been dissolved.

In the case of the problem of other minds, it is granted that all we can observe of other people is their physical behaviour (and, for an audience some sixty or seventy years later, electrochemistry in their central nervous systems). The problem is whether they really have minds and thoughts, or are just mindless chemical machines. But it is granted that both these alternatives would issue in exactly the same observations of others, and, by empiricism, such observation exhausts the evidence for either. Hence, the alternatives do not differ in evidence and so do not differ at all. There is thus no choice between them, and so no problem of other minds.

Perhaps the truly devout will hang on to their faith no matter what. In that case it looks as though nothing could count against it. Equally, no experience really counts as evidence for faith. But if faith is free from sensory evidence, then, since by verificationism meaning *is* evidence, religion is meaningless. It was not an accident that believers found such irreligion (and affronts to ethics and poetry as well) in Ayer's potboiler, *Language, Truth and Logic*. (One thing that is distressing here is that some thought they could defend their faith only by showing it empirically verifiable rather than by rebutting verificationism. The devil has no right to pick the battlefield.)

4 Objections to verificationism

One sort of objection to verificationism begins with the observation that evidence for a present claim about ancient events will be contemporary with the claim rather than with the events. Then the claim about the past seems paradoxically to be about the present, and this paradox seems to be an objection to verificationism. But the objection seems to be founded in a confusion between the date of a report and the date, if any, mentioned by the report. It may be mysterious how words, simultaneous with their present meanings, can reach across time to refer to things past (see Proper names §4). But we have yet to be shown how, if present meaning is present evidence, that mystery is any deeper. Yet if present evidence must be present experience, and experience can only be experience of what is (pretty much) contemporary with it, then the objection may be more forceful. (The finiteness of the speed of light means we can see novae long past, but we cannot choose the temporal distance at which we see while we can choose that at which we refer.)

A monument in the criticism of verificationism is Carl G. Hempel's 'Problems and Changes in the Empiricist Criterion of Meaning' (1950). Let us ask in seeming innocence exactly what the verifiability theory of meaning is. It is natural to put the theory by saying that the meaning of a (declarative) sentence is the body of sensory experience that would conclusively establish the sentence; here verification is required to be knock-down. To this form of the theory, Hempel presents three objections. First, natural science, the positivists' paradigm of meaningfulness, is full of universal hypotheses, though, as the problem of induction shows, no genuinely possible body of evidence ever suffices to establish a universal claim conclusively. Second, suppose that S typifies what the positivists count as meaningful while N typifies what they count as meaningless nonsense. Note that the disjunction, S or N, follows from S; yet if a part is gibberish, one expects the whole to be so too - how can nonsense follow from sense? Third, let P be a feature one can observe an object to exhibit or to lack. One observation would suffice to establish conclusively that there is a thing with feature P. But then the sentence 'There is something with P' is meaningful, while its negation, being universal, is not, though it seems unbelievable that a sentence be meaningful, while its negation is not. These three arguments are not meant to be decisive. Instead they represent a turn in the dialectic both of Hempel's 1950 paper and of verificationism. But through the 1950s and into the 1960s, there was a rather general presumption that verificationism was dead, and that Hempel's paper was its headstone (see Logical positivism §5).

5 Recent developments

But ideas are hardy. After Hume, the gap between experience and theory grows. Perhaps statements which are just about experience can be tested one by one against experience, but the more theoretical a statement, the less it can be so tested individually. Around 1900 Pierre Duhem argued that theoretical statements can be tested against observation only in relatively large, heterogeneous bodies (see Duhem 1906; Duhem, P.M.M. §4). Half a century later, Quine (1961) inferred from verificationism that the smallest unit with meaning is the smallest unit to which empirical evidence attaches. By Duhem's thesis, such units of evidence are large and heterogeneous in the case of theoretical sentences (see Quine, W.V. §3). But that is the more general case, and in it we cannot distinguish different meanings for different sentences; as it were, mostly there are no propositions. In particular, meanings for the mathematics used in theoretical science cannot be distinguished from meanings for the natural laws used there. Thus we lose all basis for saying such mathematics is true by virtue of its meaning (analytic), while those laws are not certified by their meanings alone. Quine thus uses one central tenet of positivism (verificationism) to undercut another (the analyticity of mathematics - see Analyticity §3). This reopens a path to Platonic metaphysics in mathematics, a path positivism thought to block.

Quine develops verificationism towards a semantic holism rather alien from positivism (see Holism: mental and semantic). But from around 1970 Michael Dummett (1975), taking a cue from the later Wittgenstein (§14), revives a verificationism more reminiscent of positivism. The later Wittgenstein sometimes seems to flirt with intuitionism in the philosophy of mathematics. The strictures intuitionism would impose on classical logic hardly seem obvious. But Dummett suggests that those strictures may fall out of a proper theory of meaning (see Intuitionistic logic and antirealism §§4-6). Meaning, Wittgenstein said, is use, and use, Dummett says, is public; the upshot is what Dummett calls the extrusion of meaning (from, perhaps, the mind into, perhaps, public life). Sometimes he presents this extrusion as the proper response to two challenges, one of acquisition, another of manifestation. The point of the first is that no matter what ideas are innate, a child learning its mother tongue has only observations of which words those around it use in which circumstances on which to base its expressions of its ideas. Meaning must be publicly observable to be learned by speakers (see Private states and language). These challenges are

challenges to an account of meaning in terms of truth-conditions, an account of meaning that may have been present in Frege (1884) and that is explicit in Donald Davidson (1984). The idea here is that the meaning of a sentence is given by the conditions for its truth (see Meaning and truth §1). Taken standardly, there is no guarantee that we can tell of what is true that it is true; Dummett describes such truth as 'verification transcendent'. If the truth-conditions which constitute meaning transcend recognition, then it would seem there is no guarantee that meaning will be manifest enough to be acquired by native speakers. So, since meaning must be knowable for communication to be possible, verification-transcendent truth must be an illusion. But truth so conceived is crucial to the exposition of classical logic; when we say that a statement or its negation is always true, we mean to allow that neither need be known. If we treat from conditions of truth to those of assertibility, as Dummett thinks we must, the logic we get is, he argues, that of intuitionism.

6 Logical problems

It is a consequence of the rejection of verification-transcendent truth that all truths can be known. This may also be a consequence of verificationism itself: any truth is meaningful and so, by verificationism, has attached to it evidence that can be had, thus making the truth known. Dummett is quite clear about the commitment that all truths can be known. Here it seems important to separate possibility from knowledge. There is, presumably, a natural number n such that Caesar had exactly n hairs on his head at the instant Brutus's dagger first penetrated Caesar's body. To say that it can be known that Caesar then had n hairs may be true, while it is pretty clearly false that it is in fact known; the possibility of knowledge should allow for the fact of ignorance.

Take a system for the propositional calculus. Add to it a very weak modal logic in which necessary conditionals with necessary antecedents have necessary consequents (Gödel's axiom) and in which necessitations of theorems are also theorems; the latter is OK if all our axioms are necessary and all our other rules of inference preserve necessity. This weak modal logic is all we need about necessity and possibility on their own. As for knowledge, we make three postulates. A1 says that all truths can be known. This can be written 'If p then it is possible that (' $\diamond$') it is known that p',

$p \rightarrow \diamond Kp$.

A2 says that what is known is true. This can be written

$Kp \rightarrow p$.

A3 says that when a conjunction is known so are its conjuncts. This can be written

$K(p \,\&\, q) \rightarrow (Kp \,\&\, Kq)$.

A2 is part of all stories of knowledge, and as for A3, it is hard to see how an explicit conjunction could be known without being understood, and so its conjuncts known. Call the whole system F. In F, we can prove that all truths are known; this can be written

$p \rightarrow Kp$.

Here is how the proof goes. Suppose that p but that it is not known that p; suppose, in short, that $p \,\&\, {\sim}Kp$. By A1, $\diamond K(p \,\&\, {\sim}Kp)$. A3 says K distributes across conjunction, and weak though our modal logic is, it is strong enough to make the distribution inside the diamond. So $\diamond(Kp \,\&\, K({\sim}Kp))$. A2 says the outermost K of the second conjunct can be dropped, and weak though our modal logic is, it is strong enough to drop this K inside the diamond. So $\diamond(Kp \,\&\, {\sim}Kp)$. This says an explicit contradiction could be true, a possibility refutable even in our weak modal logic. So our original supposition that $p \,\&\, {\sim}Kp$ has been reduced to absurdity, and thus we have shown that $p \rightarrow Kp$. (This argument has a curious provenance. It was published by Fitch in 1964, but Fitch attributes it to the anonymous referee of a paper he submitted, but did not publish, in 1945.)

General omniscience is utterly preposterous. If nothing else, it would collapse the distinction between knowledge and truth, leaving nothing to discover. If not all truths are known, then at least one of the postulates of F must be mistaken. Faulting the propositional calculus here seems excessive - like using cannon to hunt sparrows. If one is willing to countenance modality at all, then Gödel's axiom seems acceptable. Whether the rule of necessitation RN is acceptable depends on whether the axioms of F are necessary and the other rules preserve necessity. It is

generally held by those who take necessity seriously that the laws of the propositional calculus are necessary and that, say, *modus ponens* preserves necessity. Gödel's axiom too is generally thought necessary by the friends of modality. That leaves only our axioms A1-A3 of knowledge. A2 says that only truths are known; this claim is generally held to be analytic and thus necessarily true. A3 says that the conjuncts of a known conjunction are known. This claim follows, for example, from the more general claim that the logical consequences of what is known are also known. The more general claim is usually rejected since, for example, one seems to be able to know the axioms of number theory but not to know all its theorems. Nevertheless, it is only A3 and not the more general claim that is at issue here. It is at least difficult to imagine how an explicit conjunction might be known, and hence understood, unless its conjuncts also were known. If so, A3 is necessarily true and, thus, true. That leaves only A1, the focus of our original interest. If A1 is not true, then neither is it necessarily true, in which case RN is not to be trusted in *F*. Quite independently of *F*, there is a suspicious hubris in A1. What would give sentient beings the arrogance to suppose themselves so smart as to be able to figure out everything? Perhaps good manners as well as good judgment urge the denial of A1.

Still, one might feel as if a trick were being played on one; can just this dinky little logic chopping really be all it takes to show there absolutely must be ignorance? This seems like too much out of too little to be believed. So one might think that maybe there is a blunder somewhere, and one thing in the argument that bothers some people is the way it nests K inside itself. So one might consider typing (assigning 'levels' or 'types' to) knowledge, rather as Russell (1908) typed classes, or Tarski (1933) levelled languages. It is by no means obvious how to distinguish types, but custom suggests treating them like numbers. Then we would have not a single K, but rather an infinity K_1, K_2 and so on *ad infinitum*. Aping Russell, our syntax might insist that K_n be applicable to a formula *p* only if, when there are operators K_m in *p*, *n* is the successor of the largest such *m*. A2 and A3 could be postulated for each such operator uniformly. The novelty would crop up in A1 to block the first step in the proof that $p \to K_n p$. Moreover, as Hans Kamp and Charles Parsons showed independently, the result is a consistent system in which $p \to K_n p$ fails.

This result depends on sealing off the positions in which types are mentioned from quantification. For otherwise we could re-introduce K so that K*p* means that $(\exists n)K_n p$, and in this way recover the old proof that $p \to Kp$. We are thus landed in an indefinitely extended realm of types (or levels or contexts) that we have to be able to distinguish and mention. Presumably we could introduce a predicate true of all and only the types. This makes it seem like mere artifice or arbitrary convention (like not using one's hands in soccer) not to allow quantification over types. We cannot even say that knowledge is always of some type. In other words, typing without quantification is more like hiding the paradox than solving it. There is nothing special here about knowledge, for types without quantification also seem equally unsatisfactory solutions to the set-theoretic and semantic paradoxes.

Typing is not the only way one might try to diagnose trickery in *F*. A more radical move would be to take our results about *F* as a refutation of A1, as a proof that not all truths can be known and so as a refutation of the verifiability theory of meaning. What might the falsity of verificationism tell us about meaning? Verificationism would fail if there were differences in meaning which no differences in evidence - and thus, granted empiricism, sensory evidence - answered. That would in turn happen were there two sentences that differed in meaning, which would be most convincing if they could or did differ in truth-value, but such that there would be no possible way to tell by looking, listening or otherwise observing, which is true. Here we have reached a description of Descartes' problem of the external world. The matter hypothesis says that there is matter, while the dreaming hypothesis does not. On the very face of it, the two hypotheses are incompatible, and so not synonymous. If a language is to be usable for communication among its speakers, then they must know what its expressions mean. In general, when two of its expressions seem non-synonymous to its speakers, they will be non-synonymous. The dreaming and matter hypotheses seem non-synonymous to us. So if the verifiability theory of meaning says they are synonymous, then so much the worse for verificationism. As a response to the problem of the external world, verificationism was whistling in the dark.

Differences in meaning are not guaranteed to be rooted in sensory differences. Put otherwise, in a more archaic vocabulary, ideas are not images; nor are they faint copies of impressions. This is hardly hot news, and yet it raises in an acute way a worry about whether an empirically acceptable account of what meaning is can be had. What kind of a thing is a meaning? How can that be a difficult question given that since you and I know a natural language, we are both familiar with lots of meanings?

See also: Logical and mathematical terms, glossary of

W.D. HART

References and further reading

Ayer, A.J. (1936) *Language, Truth and Logic*, London: Gollancz, 2nd edn, 1946. (Widely read and highly provocative. The second edition includes a famous introduction in which Ayer wrestles with objections to the verification principle.)

Carnap, R. (1928) *Scheinprobleme in der Philosophie*, Berlin; trans. R.A. George, *Pseudoproblems in Philosophy*, with *The Logical Structure of the World*, Berkeley, CA: University of California Press, 1967.(Two seminal works by the leading member of the Vienna Circle.)

Craig, E.J. (1982) 'Meaning, Use and Privacy', *Mind* 91: 541-64.(The first section of this paper is a critical discussion of Moritz Schlick's argument in favour of the verification principle.)

Davidson, D. (1984) 'Truth and Meaning', in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 17-36.(A classic statement of the view that, taken properly, reference fixes sense.)

Descartes, R. (1641) *Meditationes de prima philosophia*, trans., 'Meditations on First Philosophy', in *Descartes: Philosophical Writings*, trans. and ed. G.E.M. Anscombe and P.T. Geach, Indianapolis, IN: Bobbs-Merrill, 1971.(The book that started modern classical philosophy.)

Duhem, P. (1906) *La Théorie physique, son objet et sa structure*, Paris: Chevalier et Rivière, 2nd edn, 1913; trans.P.P. Wiener, *The Aim and Structure of Physical Theory*, Princeton, NJ: Princeton University Press, 1992. (The source of the view that even all possible experience fails to settle all theoretical questions about the world.)

Dummett, M. (1975) 'The Philosophical Basis of Intuitionistic Logic', *Studies in Logic and the Foundations of Mathematics* 80: 5-40; repr. in *Truth and Other Enigmas*, London: Duckworth, 1978.(An authoritative statement of Dummett's contemporary version of 'meaning-empiricism'. No easy read, but the most interesting defence of intuitionism in print.)

Edgington, D. (1985) 'The Paradox of Knowability', *Mind* 94: 557-68.(Discusses issues raised in §6 of this entry.)

Fitch, F.B. (1964) 'A Logical Analysis of Some Value Concepts', *Journal of Symbolic Logic* 28: 135-42.(The first publication of the proof that if all truths can be known, then all truths are in fact known.)

Frege, G. (1884) *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: Koebner; trans. J.L. Austin, *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell, 2nd edn, 1980, esp. Introduction.(Still the best single book in philosophy of mathematics ever written.)

Hempel, C.G. (1950) 'Problems and Changes in the Empiricist Criterion of Meaning', *Revue internationale de philosophie* 4: 41-63; repr. in L. Linsky (ed.) *Semantics and the Philosophy of Language*, Urbana, IL: University of Illinois Press, 1952.(Highly influential critique of logical positivism.)

Hume, D. (1739-40) *A Treatise of Human Nature*, ed. D.G.C. MacNabb, London: Fontana, 1962.(A masterpiece. See book 1 for Hume's sceptical account of meaning.)

Kant, I. (1781/1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1963.(Another masterpiece. The second edition of 1787 contains the discussion outlined in §1.)

Quine, W.V. (1961) 'Two Dogmas of Empiricism', in *From a Logical Point of View: Nine Logico-Philosophical Essays*, Cambridge, MA: Harvard University Press, 2nd edn, 1980.(Proposes a revamped meaning-empiricism of a holistic rather than molecular or atomistic sort.)

Russell, B.A.W. (1908) 'Mathematical Logic as based on the Theory of Types', *American Journal of Mathematics* 30: 222-62; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992, 59-102.(Perhaps Russell's single best piece of work in logic.)

Tarski, A. (1933) Pojęcie prawdy w językach nauk dedukcyjnych, Warsaw; trans. J.H. Woodger (1956), 'The Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(The start of modern theories of truth.)

Williamson, T. (1982, 1987) 'Intuitionism Disproved?', *Analysis* 42: 203-8, 'On Knowledge and the Unknowable', *Analysis* 47: 154-8, and 'On the Paradox of Knowability', *Mind* 96: 256-61.(Articles which discuss logical issues of the kind described in §6 of this entry.)

Meaning in Islamic philosophy

The discussion of the notion of meaning in Islamic philosophy is heavily influenced by theological and legal debates about the interpretation of Islam, and about who has the right to pronounce on interpretation. The introduction of Greek philosophy into the Islamic world produced a new set of authorities on how to interpret texts, and this led to arguments over the potential benefits of the new approaches as compared with the traditional Islamic sciences. The discussion came to centre on the nature of ambiguity, equivocation and analogy, with different philosophers adopting diverse theories and thus attaining a variety of conclusions about how to interpret meaning. These variations have powerful implications for the understanding of their thought. Not only do the different approaches result in different conclusions, they also represent different approaches to the whole philosophical enterprise. The topic of meaning is not so much an aspect of Islamic philosophy as an interpretation of how to do Islamic philosophy itself. The main issues focus on identifying the people best qualified to interpret texts, valid interpretations of the texts, and the notion of meaning that should be employed in our understanding of the texts.

1 Religious context

Since the Qur'an was transmitted in Arabic, an understanding of the nature of that language is a vital aspect to an understanding of the text itself. Those brought up within the traditions of jurisprudence, grammar and theology were of the opinion that they were the best positioned to pronounce on the meaning of the text (see Islamic theology §1). All scriptures require interpretation, and a wide diversity of views arose within the Islamic community over the correct reading of much of the Qur'an, with the creation of different schools of thought based upon political and religious divisions such as those between the Sunni and the Shi'i communities and between the Ash'arites and the Mu'tazilites (see Ash'ariyya and Mu'tazila §1). It was accepted quite early on that while some parts of the Qur'an are clear, others are less easy to grasp and so require more complex interpretations. Some passages are *zahir* (exoteric) while others are *batin* (esoteric), and naturally the commentators disagreed on occasion over which texts fell into which category and how the esoteric texts were to be read. In a religious text, of course, even the apparently plain and commonplace can be given a richer and deeper interpretation, and different interpreters produced different interpretations. None of these was arbitrary. All were based upon argument and a variety of Islamic sources, but this hermeneutic process was capable of arriving at a variety of conclusions.

2 Language versus logic

The arrival of Greek logic in the Islamic world caused great controversy in the fourth century AH/tenth century AD. It brought with it the view that logic is superior to language since the latter deals only with the contingent and conventional, while logic encompasses the necessary and the universal. Language deals only with *alfaz* (utterances), while logic gets to the heart of the matter by analysing *ma'ani* (concepts) themselves. This controversy was explored in the celebrated debate between the grammarian Abu Sa'id al-Sirafi and the logician Abu Bishr Matta in Baghdad in AH 320/AD 932. Matta argued that logic was more important than language since the meanings which are embedded in a particular language are analysable without reference to that particular language. Those meanings could exist in a whole variety of languages. Logic is the only rigorous tool for judging when language is used correctly or otherwise, and the logician is the best qualified to adjudicate on such issues.

This was a very important debate, since if Matta were to carry the day it would imply that the traditional approach to Islamic texts rests on an error. Only logic, as understood by the Greeks and a non-Muslim such as Matta, is a sound vehicle for the understanding of texts. Al-Sirafi was clear on the significance of the debate, and he argued that it is impossible to separate language and logic in the way Matta wants. Although logic is useful at one level in dealing with concepts, it is far from comprehensive and is better suited to analysing the Greek language than the Arabic. Problems in understanding Arabic texts can only be answered by a good understanding of the Arabic language and the culture which surrounds it. Logic by itself is not sufficient (see Logic in Islamic philosophy).

This sort of debate continued in the Islamic world for some time, albeit in a more sophisticated form. The debate is highly significant, since it is really about the appropriate notion of meaning to be employed. Must that notion of meaning be taken from the context in which the text to be analysed has itself originated, or can it come from elsewhere? How one answers this question has radical implications for the way in which Islamic philosophy is to

be pursued, and the protagonists of the argument are well aware of this.

3 Ambiguity, equivocation and analogy

There are two main theories of meaning in Islamic philosophy, one broadly Neoplatonic and adopted by Ibn Sina (Avicenna) and al-Ghazali, and one broadly Aristotelian and defended by Ibn Rushd (Averroes). On the former account, the definition of *x* does not include the existence of *x*, so that something else is required to move *x* from the realm of possibility to the realm of actuality if it is to be instantiated. For Ibn Sina the mover is another thing, ultimately God, which causally necessitates the change, while for al-Ghazali it is God who directly established quite arbitrary rules for the behaviour of contingent phenomena. This means that for al-Ghazali it is possible to think of something happening without its customary explanation, that is, without a natural explanation. God could always have brought it about that people on occasion write philosophy books without possessing a head, or that straw does not burn when in contact with fire.

Ibn Rushd operates with a different concept of meaning, in accordance with which the cause of an event is part of its definition or essence. If headless people were to start composing philosophical books we should need to construct an entirely novel conceptual scheme for such a possibility to make sense. Al-Ghazali, by contrast, argues that we can conceive of such changes to our conceptual scheme by using our imagination, and failure to approve the possibility of such changes is merely a result of intellectual laziness, not the flouting of necessity. It is possible, then, for God to do anything he wishes so long as it is logically possible. God can resurrect the dead, he can intervene in the natural world, he can be aware of events in the world, he can create that world at any time he wishes out of nothing. These are all actions which Ibn Rushd argued are not possible even for God, if what is meant when talking about such actions is similar to what we mean when we talk about what we do. Al-Ghazali suggests that any analysis of the properties of God which interprets them as equivocal, ambiguous or metaphorical is a subtle attack upon the notion of God in its religious sense. It involves pretending to make God part of one's metaphysics but changing the way in which one talks about him to such an extent that he no longer is equivalent to the God of Islam.

For Ibn Rushd, there are serious problems in talking about God using the same sort of language we use about ourselves. God cannot be defined in terms of a genus or species since he cannot consist of a plurality of qualities as we do. Rather, he is the exemplar of all things, and we must work towards a conception of him by thinking analogically about the things in the world of which we have experience. The way in which al-Ghazali talks about God seems to be in line with religion, but really it involves treating the deity as someone much like ourselves but more so, and as a plurality of predicates rather than a complete unity. Ibn Rushd goes along with Aristotle's argument that there can be no priority or posteriority within the same genus (see Aristotle §§7-8), and so develops a theory of meaning which is based upon the *pros hen* notion of equivocation rather than the genus-species relation. We can use the same language about God as we use about ourselves, but we should realize that the latter type of use is derivative upon the former, since the latter concepts are aspects of the paradigm which is to be found when we talk about God. When we use the same concept to refer to God and to ourselves we are speaking not univocally but equivocally, acknowledging the very real difference which exists between the level of the human and the level of the divine.

4 The main debate

In al-Ghazali's concept of meaning, the appropriate people to interpret religious texts are those professionally involved in the religion, such as theologians, jurisprudents and so on. They need to understand the religious context of the text and apply the terms univocally to God and to his creatures. The correct way to think analytically is to adopt the methodology of the thought-experiment, holding ideas together in the imagination to see if they can be combined without contradiction. If we can imagine dead people being resurrected and leaving their graves to continue a physical existence somewhere else, in hell or paradise, then there is nothing impossible about that idea and there is no need to suggest that it is a metaphorical or equivocal reference to something else. If the Qur'an refers to physical resurrection, and if we can think about physical resurrection without contradiction, then why not just accept that what is meant by physical resurrection is what we would normally understand by that miraculous event?

It follows from Ibn Rushd's approach that the best people to interpret difficult theological passages are not the

theologians but the philosophers, since only the latter are skilled in understanding analytical and demonstrative thought. The philosopher can understand how a religious text may embody a whole range of meanings, some intended for a more sophisticated audience and some designed for a more naïve and practically-oriented audience. The latter might understand by physical resurrection that the consequences of what we do outlive us, and the moral status of what we do affects us in this life and others after we are dead. They might find this easier to understand if they can think of themselves, or something like themselves, surviving their death. The more sophisticated audience would understand by the religious text what the philosophers understand by it, what it really means. This sort of audience does not require the rhetorical and poetic language which is capable of moving the largest section of the community to action. Only philosophers are capable of resolving the meaning of a text once and for all. The other theoreticians who work within the Islamic sciences, such as theology and jurisprudence, will constantly disagree about the appropriate meaning, and as a result they will sow discord in the community. The role of the philosopher is to resolve the meaning of the text and then communicate that meaning to different kinds of audiences in different kinds of ways, each suited to the limitations of the audience to understand the real point of the text.

It can thus be seen that the issue of meaning in Islamic philosophy concerns not only the philosophy of language but also politics. It is linked to the question as to who is entitled to derive the meanings of a text and how that meaning may be communicated to others. Although the focus of discussion is generally on the relationship between ordinary language and language about God, it has far broader implications, and leads to a diversity of views on how to do philosophy itself.

See also: Aesthetics in Islamic philosophy; Epistemology in Islamic philosophy; al-Ghazali; Ibn Rushd; Ibn Sina; Logic in Islamic philosophy; Political philosophy in classical Islam

OLIVER LEAMAN

References and further reading

Abed, A. (1991) *Aristotelian Logic and the Arabic Language in Alfarabi*, Albany, NY: State University of New York Press.(A detailed and scholarly investigation of the concept of meaning and the logic versus language debate.)

Abed, S. (1996) 'Language', in S.H. Nasr and O. Leaman (eds) *History of Islamic Philosophy*, London: Routledge, ch. 52, 898-925.(Interesting and detailed account of the role that language plays in Islamic philosophy.)

Ibn Rushd (1179-80) *Fasl al-maqal (Decisive Treatise)*, ed. and trans. G. Hourani, *Averroes on the Harmony of Religion and Philosophy*, London: Luzac, 1976.(An outstandingly clear exposition of the basic themes of Ibn Rushd on different kinds of meaning.)

Ibn Rushd (1180) *Tahafut al-tahafut (The Incoherence of the Incoherence)*, trans. S. Van den Bergh, *Averroes' Tahafut al-tahafut*, London: Luzac, 1978.(A comprehensive comparison of the views of al-Ghazali and Ibn Rushd.)

Leaman, O. (1954-) '*Ma'na*', in *Encyclopedia of Islam*, Leiden: Brill, 347.(An account of the philosophical use of this Arabic term, equivalent in meaning to 'concept' or 'idea'.)

Leaman, O. (1988) *Averroes and His Philosophy*, Oxford: Oxford University Press; 2nd edn, 1997.(A general account of the thought of Averroes (Ibn Rushd), with the emphasis upon his approach to language.)

Leaman, O. (1997) 'Logic and Language in Islamic Philosophy', in B. Carr and I. Mahalingam (eds) *Companion Encyclopedia of Asian Philosophy*, London: Routledge, 950-64.(Summary of some of the major approaches to meaning in Islamic philosophy.)

Mahdi, M. (1970) 'Language and Logic in Classical Islam', in G.E. von Grunebaum (ed.) *Logic in Classical Islamic Culture*, Wiesbaden: Harrassowitz, 51-83.(The standard discussion of the battle between the logicians and the grammarians.)

Meaning, Indian theories of

The term artha *in Sanskrit is used for the notion of meaning, in the widest sense of the word 'meaning'; it can be the meaning of words, sentences and scriptures, as well as of nonlinguistic gestures and signs. Its meaning ranges from a real object in the external world referred to by a word to a mere concept of an object which may or may not correspond to anything in the external world. The differences regarding what 'meaning' is are argued out by the philosophical schools of Nyāya, Vaiśeṣika, Mīmāṃsā, Buddhism, Sanskrit grammar and Sanskrit poetics. Among these, Nyāya, Vaiśeṣika and Mīmāṃsā have realistic ontologies. Mīmāṃsā focuses mainly on interpreting the Vedic scriptures. Buddhist thinkers generally depict language as giving a false picture of reality. Sanskrit grammar is more interested in language than in ontology, while Sanskrit poetics focuses on the poetic dimensions of meaning.*

Generally, the notion of meaning is stratified into three or four types. First there is the primary meaning. If this is inappropriate in a given context, then one moves to a secondary meaning, an extension of the primary meaning. Beyond this is the suggested meaning, which may or may not be the same as the meaning intended by the speaker. Specific conditions under which these different varieties are understood are discussed by the schools.

The various Indian theories of meaning are closely related to the overall stances taken by the different schools. Among the factors which influence the notion of meaning are the ontological and epistemological views of a school, its views regarding the role of God and scripture, its focus on a certain type of discourse, and its ultimate purpose in theorizing.

1 *Artha* in different Indian traditions

The most common Sanskrit term for meaning is *artha*. In the Western literature on the notion of meaning in the Indian tradition, various terms, such as 'sense', 'reference', 'denotation', 'connotation', 'designatum' and 'intension', have been used to render the Sanskrit. However, these terms carry specific nuances of their own, and no single term adequately conveys the idea of *artha*. *Artha* basically refers to the object signified by a word. In numerous contexts, it stands for an object in the sense of an element of external reality. For instance, Patañjali (second century BC) says that when a word is pronounced, an *artha*, 'object', is understood. So in the case of 'Bring a bull' and 'Eat yoghurt', it is the *artha* that is brought in and the *artha* that is eaten.

The logicians and ontologists belonging to the schools of Nyāya and Vaiśeṣika, and the later combined school of Nyāya-Vaiśeṣika, set up an ontology containing substances, qualities, actions, relations, generic and particular properties, and absences (see Nyāya-Vaiśeṣika §§4-5). With this realistic ontology in mind, they argue that if the relation between a word and its *artha* were a natural ontological relation, there should be real experiences of burning and cutting in one's mouth after hearing words like *agni* ('fire') and *asi* ('sword'). Therefore this relation must be a conventional one (*saṃketa*), the convention being established by God as part of his initial acts of creation. The relationship between a word and the object it refers to is thought to be the desire of God that such-and-such a word should refer to such-and-such an object. It is through this established conventional relationship that a word reminds the listener of its meaning.

The school of Mīmāṃsā represents the tradition of the exegesis of the scriptural Vedic texts (see Mīmāṃsā). However, in the course of discussing and perfecting principles of interpretation, this school developed a full-scale theory of ontology and an important theory of meaning. For the Mīmāṃsakas, the primary tenets are that the Vedic scriptural texts are eternal and uncreated, and that they are meaningful. In this orthodox system, which remarkably defends the scriptures but dispenses with the notion of God, the relationship between a word and its meaning is innate and eternal. Both Nyāya-Vaiśeṣikas and Mīmāṃsakas regard language as referring to external states of the world and not just to conceptual constructions.

While the various schools of Buddhism differ among themselves concerning the nature of the external world, they all seem to agree that language relates only to a level of conceptual constructions which have no direct relationship to the actual state of the world. The tradition represented by Theravāda and the Vaibhāṣikas argued that a word refers to a thing which, in reality, is nothing but a composite entity made up of components which are momentary and in a continual flux (see Momentariness, Buddhist doctrine of). The components, the momentary atomic

elements (*dharma*), are presumed to be more real, but words do not refer to this level of reality. Thus language gives us a less than true picture of what is out there. Other schools of Buddhism, such as Vijñānavāda, reduced everything to fleeting states of consciousness (*vijñāna*). From this point of view, the objects referred to by words are not even composites. They are more like fictions (*vikalpa*) or illusions (*māyā*) created by a magician. The Mādhyamika school of Buddhism focused on the essential emptiness (*śūnyatā*) of all objects which are subject to dependent origination (*pratītyasamutpāda*). This also leaves language far away from the level of reality, which in this case is the level of emptiness (see Buddhist concept of Emptiness). Later Buddhist logicians, such as Dignāga, developed a theory of word-meaning which we can call the 'exclusionary' theory of meaning (*apoha*). Briefly stated, if a word refers only to a conceptual construction, and not to a state of reality, then how are we to construe the conceptual construction? The *apoha* theory proposed by the Buddhist logicians says that external reality ultimately consists of momentary atomic elements which are so individualized and unique (*svalakṣaṇa*) that they are beyond perception and characterization. Our perceptions and conceptions involve generalization (*sāmānyalakṣaṇa*), and hence do not correspond to reality. A concept which corresponds to a given word must be finally construed as being nothing more than the exclusion of all other concepts. This theory ultimately says that all concepts are different from each other, and yet they cannot be defined by making a reference to any level of reality (see Nominalism, Buddhist doctrine of).

The tradition of the grammarians, beginning with Bhartṛhari (fifth century), seems to have followed a middle path between the realistic theories of reference (*bāhyārthavāda*) developed by Nyāya-Vaiśeṣika and Mīmāṃsā on the one hand, and the notional/conceptual theories (*vikalpa*) of the Buddhists on the other. For them, the meaning of a word is closely related to the level of understanding. Whether or not things are real, we do have concepts. These concepts form the content of a person's cognitions derived from language. Without necessarily denying the external reality of objects in the world, the grammarians claimed that the meaning of a word is only a projection of intellect (*bauddhārtha*, *buddhipratibhāsa*). The examples they offer, such as *śaśasṛōga* ('horn of a rabbit') and *vandhyāsuta* ('son of a barren woman'), remain meaningful within this theory. The Sanskrit grammarians are thus not concerned with the truth-functional value of linguistic expressions. For them, the truth of an expression and its meaningfulness are not to be equated.

2 Varieties of meaning

By the middle of the second millennium AD, a certain uniformity came about in the technical terminology used by different schools. The prominent schools in this period were the new school of Nyāya - Navya-Nyāya - initiated by Gaōgeśa, and the schools of Mīmāṃsā, Vedānta, and Sanskrit grammar. While all these schools engaged in pitched battles against each other, they seem to have accepted the terminological lead of the neo-logicians, the Navya-Naiyāyikas. Following the discussion of *artha* by the neo-logician Gadādhara, we can state the general framework of a semantic theory. Other schools accepted this general terminology, with some variations.

It may be said that the term *artha* stands for the object or content of a verbal cognition or a cognition which results from hearing a word (*śābdabodhaviṣaya*). Such a verbal cognition results from the cognition of a word (*śābdajñāna*) on the basis of an awareness of the signification function pertaining to that word (*padaniṣṭhavṛttijñāna*). Depending upon the kind of signification function (*vṛtti*) involved in the emergence of the verbal cognition, the meaning belongs to a distinct type. In general terms:

1 When a verbal cognition results from the primary signification function (*śakti/abhidhāvṛtti/mukhyavṛtti*) of a word, the object or content of that cognition is called primary meaning (*śakyārtha/vācyārtha/abhidheya*).
2 When a verbal cognition results from the secondary signification function (*lakṣaṇāvṛtti/guṇavṛtti*) of a word, the object or content of that cognition is called secondary meaning (*lakṣyārtha*).
3 When a verbal cognition results from the suggestive signification function (*vyañjanāvṛtti*) of a word, the object or content of that cognition is called suggested meaning (*vyōgyārtha/dhvanitārtha*).
4 When a verbal cognition results from the intentional signification function (*tātparyavṛtti*) of a word, the object or content of that cognition is called intended meaning (*tātparyārtha*).

Not all schools of Indian philosophy accept all of these different kinds of signification functions for words, and they hold substantially different views on the nature of words, meanings, and the relations between words and meanings. However, the above terminology holds true, in general, for most of the schools. Let us note some of the

important differences. Mīmāṃsā claims that the sole primary meaning of the word 'bull' is the generic or class property (*jāti*) - 'bullness', say - and the individual object which possesses this generic property, namely a particular bull, is only secondarily and subsequently understood from the word 'bull'. The school called Kevalavyaktivāda, a segment of the Nyāya school, argues that a particular individual bull is the sole primary meaning of the word 'bull', while the generic property bullness is merely a secondary meaning. Nyāya generally argues that the primary meaning of a word is an individual object qualified by a generic property (*jātiviśiṣṭavyakti*), both being perceived simultaneously.

Sanskrit grammarians distinguish between various different kinds of meanings (*artha*). The term *artha* stands for an external object (*vastumātra*), as well as for the object that is intended to be signified by a word (*abhidheya*). The latter - that is, meaning in a linguistic sense - can be meaning in a technical context (*śāstrīya*), such as the meaning of an affix or a stem, or it can be meaning as understood by people in actual communication (*laukika*). Then there is a further distinction. Meaning may be something directly intended to be signified by an expression (*abhidheya*), or it can be something which is inevitably signified (*nāntarīyaka*) when something else is really the intended meaning. Everything that is understood from a word on the basis of some kind of signification function (*vṛtti*) is covered by the term *artha*. Different systems of Indian philosophy differ from each other over whether a given cognition is derived from a word on the basis of a signification function, through inference (*anumāna*), or through presumption (*arthāpatti*). If a particular item of information is deemed to have been derived through inference or presumption, it is not included in the notion of word-meaning.

3 Other dimensions of *artha*

The scope of *artha* is actually not limited in Sanskrit texts to what is usually understood as the domain of semantics in the Western literature. It covers elements such as gender (*liṅga*) and number (*saṃkhyā*), as well as such semantic/syntactic roles (*kāraka*) as 'agentness' (*kartṛtva*) and 'objectness' (*karmatva*). Tenses such as the present, past and future, and moods such as the imperative and optative are also traditionally included in the *arthas* signified by a verb root or an affix.

Another aspect of the concept of *artha* is revealed in the theory of *dyotyārtha*, 'co-signified' meaning. According to this theory, to put it in simple terms, particles such as *ca* ('and') do not have any lexical or primary meaning. They are said to help other words used in constructions with them to signify special aspects of their meaning. For instance, in the phrase 'John and Tom', the meaning of grouping is said to be not directly signified by the word 'and'. The theory of *dyotyārtha* argues that grouping is a specific meaning of the two words 'John' and 'Tom', but that these two words are unable to signify it if used by themselves. The word 'and' used along with them is said to work as a catalyst that enables them to signify this special meaning.

The problem of use and mention of words is also handled by Sanskrit grammarians by treating the word itself as a part of the meaning it signifies. This is a unique way of handling this problem.

4 Different views regarding sentence-meaning

Most schools of Indian philosophy have an atomistic view of meaning and the meaning-bearing linguistic unit. This means that a sentence is put together by combining words and words are put together by combining morphemic elements such as stems, roots and affixes. The same applies to meaning. The word-meaning may be viewed as a fusion of the meanings of stems, roots and affixes, and the meaning of a sentence may be viewed as a fusion of the meanings of its constituent words. Beyond this generality, different schools have specific proposals. The tradition of Prābhākara Mīmāṃsā proposes that the words of a sentence already convey contextualized/connected meanings (*anvitābhidhāna*) and that the sentence-meaning is not different from a simple addition of these inherently connected word-meanings. On the other hand, the Naiyāyikas and the Bhāṭṭa Mīmāṃsakas propose that the words of a sentence taken by themselves convey uncontextualized/unconnected meanings, and that these uncontextualized word-meanings are subsequently brought into a contextualized association with each other (*abhihitānvaya*). Therefore, sentence-meaning is different from word-meanings, and is communicated through the concatenation (*saṃsarga*) of words rather than by the words themselves. This is also the view of the early grammarians such as Patañjali and Kātyāyana.

For the later grammarian-philosopher Bhartṛhari, however, there are no divisions in speech acts and in

communicated meanings. He says that only a person ignorant of the real nature of language believes the divisions of sentences into words, stems, roots and affixes to be real. Such divisions are useful fictions and have an explanatory value in grammatical theory, but have no reality in communication. In reality, there is no sequence in the cognitions of these different components. The sentence-meaning becomes an object or content of a single flash of cognition (*pratibhā*).

5 Some important conceptions

The terms *śakyatāvacchedaka* and *pravṛttinimitta* signify a property which determines the inclusion of a particular instance within the class of possible entities referred to by a word. It is a property whose possession by an entity is the necessary and sufficient condition for a given word to be used to refer to that entity. Thus the property of potness may be viewed as the *śakyatāvacchedaka* controlling the use of the word 'pot'.

Lakṣaṇāvṛtti (the 'secondary signification function') is invoked in situations where the primary meaning of an utterance does not appear to make sense in view of the intention behind the utterance, causing one to look for a secondary meaning. However, the secondary meaning is always something that is related to the primary meaning in some way. For example, the expression *gaṅgāyāṃ ghoṣaḥ* literally refers to a colony of cowherds on the river Ganges. Here it is argued that one obviously cannot have a colony of cowherds sitting on top of the Ganges. This would clearly go against the intention of the speaker. Thus there is both a difficulty in justifying the linkage of word-meanings (*anvayānupapatti*) and a difficulty in justifying the literal or primary meaning in relation to the intention of the speaker (*tātparyānupapatti*). These interpretive difficulties nudge one away from the primary meaning of the expression to a secondary meaning which is related to that primary meaning. Thus we understand the expression as referring to a colony of cowherds on the *bank* of the Ganges.

It is the third level of meaning, or *vyañjanāvṛtti* ('suggestive signification function'), which is analysed and elaborated by authors such as Ānandavardhana (ninth century) in the tradition of Sanskrit poetics. Consider the following instance of poetic suggestion. Her husband away on a long journey, a lovelorn young wife instructs a visiting young man: 'My dear guest, I sleep here and my nightblind mother-in-law sleeps over there. Please make sure you do not stumble at night.' The suggested meaning is an invitation to the young man to come and share her bed. Thus the poetic aspect of language goes well beyond the levels of lexical and metaphorical meanings, and heightens aesthetic pleasure through such suggestions.

6 Why the differences?

The nuances of these different theories are closely related to the markedly different interests of the schools within which they developed. Scholars of Sanskrit poetics were interested in the poetic dimensions of meaning. Grammarians were interested in language and cognition, but had little interest in ontology. For them, words and meanings had to be explained irrespective of one's metaphysical views. Nyāya-Vaiśeṣikas were primarily interested in logic, epistemology and ontology, and argued that a valid sentence was a true picture of a state of reality. The foremost goal of Mīmāṃsā was to interpret and defend the Vedic scriptures. Thus, for Mīmāṃsā, meaning had to be eternal, uncreated and unrelated to a person's intention, because its word *par excellence*, the Vedic scriptures, was eternal, uncreated and beyond any authorship, divine or human. The scriptural word was there to instruct people on how to perform proper ritual and moral duties. The Buddhists, on the other hand, aimed at weaning people away from all attachment to the world, and hence at showing the emptiness of everything, including language. They were more interested in demonstrating how language fails to portray reality, than in explaining how it works. The theories of meaning were thus a significant part of the total agenda of each school and need to be understood in their specific contexts.

See also: Interpretation, Indian theories of; Language, Indian theories of; Language, philosophy of; Meaning in Islamic philosophy

MADHAV M. DESHPANDE

References and further reading

Biardeau, M. (1967) *Théorie de la connaissance et philosophie de la parole dans le brahmanisme classique (Theory of the Understanding and Philosophy of the Word in Classical Brahmanism)*, Paris: Mouton.(A comprehensive discussion of theories of meaning and their evolution and interaction in the traditions of

Sanskrit grammar, Mīmāṃsāand Nyāya-Vaiśeṣika; no English translation is available as yet.)

Deshpande, M.M. (1992) *The Meaning of Nouns: Semantic Theory in Classical and Medieval India*, Dordrecht: Kluwer Academic Publishers.(The bulk of the book is an annotated translation of a seventeenth-century Sanskrit text on the meaning of nouns. The introduction covers the history of a number of semantic theories in Sanskrit grammar and philosophy.)

Matilal, B.K. (1971) *Epistemology, Logic and Grammar in Indian Philosophical Analysis*, The Hague/Paris: Mouton.(Very readable general introduction to Indian approaches to meaning, logic and language.)

Matilal, B.K. (1985) *Logic, Language and Reality: An Introduction to Indian Philosophical Studies*, Delhi: Motilal Banarsidass.(More extensive, but still very accessible, discussion of a wide range of issues in Indian philosophy of language.)

Raja, K. (1963) *Indian Theories of Meaning*, Adyar Library Series 91, Madras: Adyar Library and Research Centre.(A somewhat dated, but still very useful, account.)

Metaphor

A standard dictionary definition describes a metaphor as 'a figure of speech in which a word or phrase literally denoting one kind of object is used in place of another to suggest a likeness between them'. Although the theoretical adequacy of this definition may be questioned, it conveys the standard view that there is a difference between literal and nonliteral language; that figurative speech is nonliteral language and that a metaphor is an instance of figurative speech.

The three most influential treatments of metaphor are the comparison, interaction and speech act theories. According to the first, every metaphor involves a comparison; a specific version of this view is that every metaphor is an abbreviated simile. According to the second, every metaphor involves a semantic interaction between some object or concept that is literally denoted by some word, and some concept metaphorically predicated on that word. According to the third, it is not words or sentences that are metaphorical but their use in specific situations; thus, to understand how metaphors function, one must understand how people communicate with language.

1 The comparison theory

The comparison theory asserts that every metaphor compares two things, one of which is designated by a word or phrase used literally and the other of which is designated by a word or phrase used metaphorically. The most interesting version of the comparison view of metaphor is the simile theory. According to its originator, Aristotle (*Rhetoric* III.4), a metaphor is an abbreviated simile: the sentence 'My lover is a red rose' is shorthand for 'My lover is like a red rose'. (The simile theory is also sometimes known as the abbreviation or substitution theory, because a metaphor is represented as being an abbreviation of or a substitution for a simile.) The merit of this view is its double simplicity. First, what looked like two things (metaphors and similes) are said to be one thing; second, the analysis that is given to similes is sometimes quite simple. A simile is true just in case there is some feature that each object has. The feature common to both constitutes the similarity.

One objection to this theory is that what has been represented as its first merit is in fact a misrepresentation. Metaphors and similes are different things because typically if a metaphor were literally asserted, the result would be a falsehood, while (virtually) all similes are true. For example, the metaphor 'My lover is a red rose' is literally false, while the simile 'My lover is like a red rose' is literally true. The pretence of saying something false is a crucial feature of how metaphors work. To hold that metaphorical sentences are in fact similes obscures this feature. This is not to deny that metaphors typically communicate a truth in some way; it is rather to hold that a metaphor communicates its truth in a different way to a simile.

Metaphors (unlike similes) do not state or express truths; they imply them. In order to speak truly, the words of a sentence must be conventionally or semantically appropriate to each other, but the words of a metaphor are not semantically appropriate in this sense. For example, in the metaphor 'Sally is a block of ice', 'Sally' and 'a block of ice' belong to categories so different that to juxtapose them results not merely in a falsehood but in an absurdity. Someone may reply that, since 'Sally is a block of ice' can be used metaphorically, 'Sally is not a block of ice' can also be used metaphorically; and since this latter metaphor results in a true assertion, its words cannot be semantically inappropriate to each other. The proper reply to this objection is to deny that 'Sally is not a block of ice' is a metaphor. (To assert it is to assert something that is obviously true, but that does not mean that the assertion is pointless any more than to assert that 'Business is business' need be pointless. To assert that 'Sally is not a block of ice' may imply that one rejects the metaphorical utterance of 'Sally is a block of ice'.) Sentences that may appear to be negations of metaphorical sentences, such as John Donne's memorable sentiment 'No man is an island', are simply instances of *meiosis* (intentionally saying something logically weaker than the speaker intends to convey, usually by using a negative). Philosophers who consider 'No man is an island' and 'Sally is not a block of ice' to be metaphors, are in effect conflating metaphor and meiosis. They make it difficult or impossible to distinguish something obviously, literally false ('Business is theft') from something obviously, literally true ('Business is business').

To deny an assertion is *ipso facto* to make an assertion; but to deny a metaphor is not to make a metaphor. Assertions are a type of speech act, performed through a literal use of words. But a metaphor is not a kind of

speech act and relies on the literal use of words only in order to indicate that the speaker means to communicate something other than their literal meaning.

Another deficiency of the standard simile theory is that it does not account for the fact that some effort or mental computation is required in order to work out what a metaphor means, in contrast with a simile which usually means what it says. In order to meet this objection, a quasi-simile theory has been proposed. Suppose someone says 'My lover is a red rose'. To understand this metaphor, one considers the set of red roses plus the speaker's lover and then tries to work out which property or properties all the members of this set share; in other words, to say that one's lover is a red rose is to say that they are similar in some respects to red roses. The objection to this theory is the same as one raised to the original theory. As Davidson (1979: 36) points out, a metaphor only intimates what a simile declares.

2 The interaction theory

According to the interaction theory, a metaphor involves an interaction between a literal element in a sentence and a metaphorical element. So there are always two parts to a metaphor. Max Black (1962) calls the literal element the frame and the metaphorical element the focus. Searle (1979) has effectively refuted this version of the theory by pointing out that some metaphorical sentences do not contain any literal element which could plausibly serve the role of frame. For example, the sentence 'The bad news congealed into a block of ice' is a mixed metaphor, but it contains no word with a literal meaning that can serve as the frame. Another objection to the interaction theory is that the term 'interaction' is itself a metaphor. Since there is no literal interaction between words in a sentence or between concepts, the interaction theory never actually explains or describes how metaphors work. More recent treatments of the theory explain interaction as the placement of the object literally denoted within a conceptual system of commonplaces associated with the metaphorical term. Thus, to say 'Mary is a wolf' is to associate Mary and her attributes with the complex of properties associated with wolves: cunning, viciousness, unpredictability and so on. None the less 'interaction' seems to be the wrong word to convey this feature. Also, the interaction theory claims that the meaning of the words is affected by this interaction and that seems to be an inaccurate description of the phenomenon.

Donald Davidson has insisted that semantically metaphors mean what the words used to convey them literally mean and nothing more. His thesis constitutes a direct rejection of all views that hold that a distinction must be drawn at the level of words or sentences between two kinds of meaning: literal meaning and metaphorical meaning. While the literal meaning is usually taken as obvious, the metaphorical meaning is often said to be ineffable and ineffability is given as the reason why metaphors are used. Sceptics about ineffability would object to a kind of meaning that is inherently mysterious or unintelligible, even if individual speakers are unable to articulate what they mean. In any case, what is needed is an explanation of how metaphors are understood. Here, Davidson seems to be right when he says that understanding a metaphor is the same kind of activity as understanding any other linguistic utterance and all understanding requires an 'inventive construal' of what the literal meaning of the metaphorical utterance is and what the speaker believes about the world. Furthermore, this activity is 'little guided by rules'. Making a metaphor, like speaking generally, is a 'creative endeavour'.

3 The speech act theory

It is fair to demand that a theory of metaphor fit into some larger theory of language or communication. While the comparison and interaction theories do not meet this demand, the speech act theory does. H.P. Grice (1989) distinguished between 'utterance occasion meaning' (the meaning that an utterance has on some particular occasion of its use) and 'timeless utterance meaning' (the meaning an utterance has based upon its usual use and independent of any particular time). This latter form of meaning can be associated with literal or conventional meaning. In a metaphor, the speaker intends to have the audience work out the 'occasion' meaning of their utterance in part by understanding its 'timeless' meaning (see Grice, H.P.).

How does this happen? Grice points out that normal conversation is directed by the cooperative principle: make your contribution to a conversation appropriate. This principle issues in various conversational maxims. For example, what a person says should be truthful and supported by evidence; it should be as informative as the audience needs it to be and not more so; and the information should be relevant to the topic of conversation. These maxims are related to the content of the message. There are also maxims concerning how the message is to be

expressed: it should be clear, unambiguous, brief and orderly. Of course these maxims are not always fulfilled. Sometimes they go unfulfilled quietly and unostentatiously, such as when a person lies or makes an innocent mistake. Sometimes they are flouted, that is, openly and ostentatiously unfulfilled in order to achieve a certain effect. For example, one way to show disgust for or disappointment with someone is to say 'You're a fine friend'. In such instances of sarcasm or irony, the speaker flouts the maxim of truthfulness and thereby within the context of the conversation means exactly the opposite of what appears to be said. Other figures of speech can be explained in similar ways. Hyperbole is intentionally making an exaggerated statement in order to achieve a certain effect. Within the theory being sketched, hyperbole is also achieved by flouting the maxim of truthfulness: the speaker makes a statement that is logically stronger than the proposition they intend to communicate.

Within this theory, meiosis is achieved by flouting the maxim of quantity (which enjoins the speaker to make their statement logically as strong as possible); for example, by saying, 'John D. Rockefeller was not poor'. The audience, trying to make sense of the speaker's apparent nonfulfilment of the cooperative principle, infers that the speaker means something stronger than was said.

Metaphor fits into the same pattern of explanation although the calculation of what a metaphor means is more complicated. If someone says 'My lover is a red rose' in circumstances that are otherwise unexceptional, it is sensible for the audience to judge that the speaker is flouting the maxim of truthfulness. What then does the speaker mean? The audience's strategy in the case of metaphor, as in every other case of linguistic understanding, is to achieve a cognitive equilibrium by accepting those propositions that make the most sense of the situation. This involves making new beliefs fit in as smoothly as possible with existing beliefs and sometimes revising old beliefs as part of this process. Briefly, for the example above, it makes sense for the audience to believe that the speaker intends the audience to recognize that the speaker intends the audience to think of certain qualities of a rose - its beauty, value, delicacy, fragrance - and then attribute them to the lover. Such a search for cognitive equilibrium is close to what Davidson meant when he said that understanding language is 'a creative endeavour'.

See also: Communication and intention; Davidson, D.; Speech acts

<div align="right">A.P. MARTINICH</div>

References and further reading

Aristotle (*c.*mid-4th century BC) *Rhetoric*, trans. G. Kennedy, *Aristotle on Rhetoric: A Theory of Civic Discourse*, New York: Oxford University Press, 1991, III 4.(The first statement of the simile theory.)

Black, M. (1962) *Models and Metaphors*, Ithaca, NY: Cornell University Press. (Contains the most influential presentation of the interaction theory.)

Davidson, D. (1979) 'What Metaphors Mean', in S. Sacks (ed.) *On Metaphor*, Chicago, IL: University of Chicago Press, 29-45.(Davidson emphasized that words themselves do not have a metaphorical meaning.)

Goodman, N. (1968) *Languages of Art*, Indianapolis, IN: Bobbs-Merrill.(A sophisticated version of the interaction theory.)

Grice, H.P. (1989) 'Logic and Conversation', in *Studies in the Way of Words*, Cambridge, MA: Harvard University Press, 22-40.(Contains Grice's general theory of conversation, including a treat ment of metaphor.)

Johnson, M. (1980) *Philosophical Perspectives on Metaphor*, Minneapolis, MN: University of Minnesota Press. (A collection of important articles on metaphor.)

Martinich, A.P. (1984) 'A Theory for Metaphor', *Journal of Literary Semantics* 13: 35-54; repr. in S. Davis (ed.) *Pragmatics*, Cambridge: Cambridge University Press, 1991, 507-18.(A speech act theory of metaphor with criticisms of Searle's version.)

Noppen, J.P. van (ed.) (1985, 1990) *Metaphor*, vol. 1, *A Bibliography of Post-1970 Publications*; vol. 2, *A Classified Bibliography of Publications*, Amsterdam: Benjamins.(Continues the bibliographical work begun by Shibles.)

Ortony, A. (ed.) (1979) *Metaphor and Thought*, Cambridge: Cambridge University Press; 2nd edn, 1994. (Contains important articles by philosophers, psychologists, linguists and literary theorists.)

Searle, J. (1979) 'Metaphor', in A. Ortony (ed.) *Metaphor and Thought*, Cambridge: Cambridge University Press; 2nd edn, 1994.(A speech act theory of metaphor in an important volume that contains contributions by philosophers, psychologists, linguists and literary theorists.)

Shibles, W. (1971) *Metaphor: An Annotated Bibliography and History*, Whitewater, WI: Language Press.(The most complete bibliography on metaphor up to 1970.)

Moscow-Tartu School

*The Moscow-Tartu School of semiotics (theory of signs) was formed when a diverse group of scholars joined informally from the 1950s to 1980s to provide alternatives to the regnant Soviet approaches to language, literature and culture. Their work develops the linguistics of Saussure, elaborated by Trubetzkoi and Hjelmslev, with its central notions of sign as union of signifier and signified, its distinction between language as system (*langue*) and language as utterance (*parole*), and its analysis in terms of the significant differences between paired equivalent elements in a system (that is, meaning is a matter not of individual elements, but of the relationship between comparable elements). In its early stages members of the Moscow-Tartu School did intricate analyses of lyric poetry and of highly conventional prose works (such as detective stories) using statistical and linguistic methods. They subsequently came to treat art works and other cultural artefacts as the products of 'secondary modelling systems', that is, as elements arranged according to rules that could be seen as language-like and hence accessible to analysis by the procedures of structuralist linguistics. The group shared an interest in Western and pre-Stalinist Russian literary theory - especially in the Russian formalists - and in contemporary linguistics, semiotics and cybernetics. In a time of pervasive intellectual stagnation this loose confederation sought to formulate objective and exact methods for literary scholarship, to republish works of Russian theory that had been repressed from the 1930s to 1950s, and to bring scholarship in the humanities into line with developments in other scholarly fields. During the 1970s prominent members of the group, such as Iu.M. Lotman and B.A. Uspenskii, turned from more theoretical and formalized work to historical studies of culture as a system of semiotic systems.*

1 History: the Soviet context and the rise of semiotics

It is difficult to comprehend the extent to which the policies of the Stalin era hindered the unfolding of academic life. A vigilant censorship, the prescriptions of party congresses, and academic hierarchies established not by intellectual distinction but by survival skills, all combined to discourage critical thinking. Literary scholarship, which had rejected the formalists' attempts to examine literature as a process with its own dynamics and evolution (see Russian literary Formalism), became particularly vulnerable to the demands of the political apparatus. By an irony of cultural history Soviet literary criticism had followed the pattern of innovation and mechanization outlined by the formalist theory of literary evolution: the innovatory and socially committed literary criticism of the nineteenth-century Russian intelligentsia, with its utilitarian view of literature and its understanding of art as a reflection of reality, had become institutionalized and stultifying, a barrier to re-examining the considerable role of literature in Russian society. Yet some disciplines, because of their remoteness, abstractness (linguistics) or technological promise (cybernetics, machine translation) were relatively exempt from the blatant ideologization of academic life. It was to these disciplines that a number of young linguists and 'philologists' (scholars trained in literature, historical linguistics and the history of ideas) turned for support in renewing the humanities. Their fascination with the 'exact methods' and 'objective scholarship' that these disciplines offered ultimately must be understood in the context of the impasse that scholarship had reached.

One of the group's most erudite members, the linguist V.V. Ivanov, would in 1976 trace the rise of Soviet semiotics to the research and theories of Russian and foreign linguists and anthropologists. He specifically identified the psychologist Vygotskii, the cinematographer Eisenstein, the information theorists A.N. Kolmogorov and Claude Shannon (see Information theory), the phenomenologist G.G. Shpet, the philologist M.M. Bakhtin and his colleague V.N. Voloshinov, for their work in reconstructing ancient cultures, in developing theories of the sign and in studying the structures and levels of art. The group's acquisition of these interests was achieved through conferences, informal gatherings and harsh debates between 'physicists' and 'lyricists' in scholarly and popular periodicals.

During the 1950s structuralist analysis slowly began to develop in the Soviet Union, enabled in part by Stalin's 1950 articles on linguistics, which in effect freed the discipline from Marxism's base-superstructure model, licensing the study of grammar and internal laws of language development. Two conferences brought the movement into focus and lent it its particular Soviet concern with cybernetics and information theory: a conference in Gorky (1961) 'On the Application of Mathematical Methods to the Study of the Languages of Artistic Literature' and a symposium in Moscow (1962) 'On the Structural Study of Sign Systems'. The first conference centred around the interest of Academician Kolmogorov in poetry's special potential for conveying information,

but already the papers showed the group's wide range of interests and approaches: I.I. Revzin used Noam Chomsky's *Syntactic Structures* (1957) to speculate on the generative structures of literary texts, Ivanov surveyed the achievements of Western structuralism, and A.K. Zholkovskii constructed a genealogy from the works of the Russian Formalists and Sergei Eisenstein. He and Iu.K. Shcheglov would subsequently develop their own 'generative poetics', a pragmatic one which focused on the effects generated by the poetic text. The Moscow Symposium, organized by Ivanov, included many of the same participants, but had a more semiotic and less cybernetic focus. Danish semiotic theory ('glossematics') (see Structuralism in linguistics §3) offered the promise of interaction between mathematics and semiotics and the possibility of joining the human sciences in a common and exact methodology. A group of papers on art as a semiotic system served as a beginning of the movement beyond verbal texts; papers on etiquette, games, and fortune-telling began the group's investigations of behaviour and popular culture.

The Tartu presence in the Moscow-Tartu School began with Iu.M. Lotman's 1963 essay 'On the Delimitation of the Concept of Structure in Linguistics and Literary Scholarship' and his *Lektsii po struktural'noi poëtike (Lectures on Structural Poetics)*, which appeared in the first issue of an important new series, *Trudy po znakovym sistemam (Works on Sign Systems)* (1964). As Uspenskii notes in his account of the School's beginnings, Tartu University brought a different scholarly culture to the movement: more oriented towards canonical literature and the history of ideas, less concerned with linguistics and with popular genres. Tartu University also brought the legacy (however constrained by Soviet reality) of having been the Russian Empire's most European university, with a tradition of academic autonomy that the other Russian universities had not always enjoyed.

The high point of the Moscow-Tartu School's activity was its five summer schools (1964-74), held at Kääriku in rural Estonia and in Tartu itself. Described by their participants as 'carnivalesque', 'utopian' and 'hermetically sealed', they centred around lively discussions rather than formal presentations, and their proceedings give brief lists of theses rather than lengthy formal papers. The summer schools brought together the Moscow semioticians and the philology faculty from Tartu, occasional guests (such as Jakobson in 1966), and a number of younger scholars who would continue to develop the school's themes and concepts after it ceased to exist.

The publications of the Moscow-Tartu School, *Trudy po znakovym sistemam* and other periodicals from Tartu, appeared in shoddy editions whose print runs could not come close to meeting the domestic and international demand for them. Their opponents in Soviet academic life published in much more widely circulated publications and levelled a consistent series of charges against the semioticians, as Peter Seyffert has noted: that semiotics was mere fashion, formalism couched in incomprehensible shibboleths; that its claim to universal applicability could not be sustained; and that 'mathematical methods' and 'exact science' were promises often unrealized and unrealizable. Certainly the School's focus on the sign and on language opposed the entrenched Russian view of art as 'thinking in images', and certainly its view of culture as a semiotic mechanism rather than a reflection of the means of production, challenged the dogma of Soviet Marxism.

For a variety of internal and external reasons, the Moscow-Tartu School had ceased to function by the mid-1970s. Growing government opposition to intellectual unorthodoxy met with opposition from the Tartu faculty, especially over the invasion of Czechoslovakia in 1968. A number of members of the School emigrated from the Soviet Union and made successful careers in the West. The growing interest of members of the School in questions of culture, which began with studies of typology and increasingly branched out into highly specific research with little attention to theory, led to methodological diffuseness. Nevertheless, the publications of the Moscow-Tartu School continued to appear in periodicals and, increasingly, in book form. By the 1990s Lotman was starring in a highly regarded television series on Russian culture. By another irony of cultural history the Moscow-Tartu School, condemned for neglecting traditional Russian views on literature and society, had become the principal interpreter of both, and of their interrelations for the new Russian Republic. After Lotman died in 1993, this canonization was completed by generous obituaries in the Russian press and by the republication of his semiotic studies.

2 Doctrine: general topics and the work of Uspenskii and Lotman

Because so many of the School's works were experimental and because their work was so far-ranging, it can be difficult to generalize about their activity. During the two decades of their most intense activity the School tried out many variants of semiotic analysis, moving from an early interest in mathematical modelling, cybernetics and

information theory to work more akin to cultural history and cultural anthropology, two disciplines which were virtually non-existent in the Soviet Union. Nevertheless, a common feature was their use of linguistic models. This was also the case with American, Danish, French and Italian structuralism, but the use of linguistic models stands out even more sharply in the Soviet context because the regnant approach to literature focused on 'images', not 'signs', and on social determinism rather than on self-regulating sign systems. Against this background the Moscow-Tartu School could be seen as a movement, even though its members were drawing upon a number of different linguistic and semiotic theories: C.S. Peirce's theory of the sign, Saussure's understanding of semiotic systems, Danish glossematics, Trubetzkoi's phonemics, Jakobson's poetry of grammar and Chomsky's transformational grammar, to name the most important (see Chomsky, N. §1; Peirce, C.S. §8; Saussure, F. de §2; Structuralism in linguistics §§2-3; Syntax §3). Yet members of the School tended not to pursue linguistic models as exhaustive sets of discovery procedures, as did Jakobson in his studies in the 'poetry of grammar': they moved from the mechanisms of meaning-production towards interpretation and beyond the study of closed systems towards contextual studies. As the group moved beyond language to other semiotic systems - 'secondary modelling systems', as Uspenskii termed them - the ties to linguistic theory could become tenuous. Nevertheless, even in these later essays one detects an enduring fascination with language, language use (pragmatics), and analytic operations learned from formal linguistics (such as analysis in terms of binary opposition). The problems that the group addressed tended to follow logically from this fundamental appeal to linguistic models: how does poetry convey information with its formal resources, how does poetry use the resources of language to model reality and the artist's worldview, how may other cultural spheres (including social behaviour) can be analysed using the model of linguistic analysis, what is the text that is available for semiotic analysis, how do different cultures in different historical periods use signs differently, what is the impact of such secondary modelling systems (such as literature, art, or the theatre) upon human cognition and behaviour? The logic of their methodology did not, however, lead them into exploring deeply such subjects as 'the death of the author' (who would become a 'guest of the text' for such French semioticians as Roland Barthes) or the theory of tropes. Unlike Western structuralists and post-structuralists, the Moscow-Tartu group tended to view the text as a unity, however multilayered and however rich its resources for generating interpretations. Language retained for them its ability to represent reality, complexities of coding notwithstanding; and the human subject was inevitably seen as the constitutor, or at least the user, of codes, more than constituted by them. In this the Moscow-Tartu semioticians, especially those who worked on such canonical figures as Pushkin and Gogol', contributed a semiotic version to the traditional Russian heroic portrayal of authors and to the traditional Russian veneration of high cultural texts.

Members of the School did innovatory work on many subjects: on various aspects of semiotic theory and on text theory, on mythology and the reconstruction of ancient symbol systems (Ivanov, V.N. Toporov, E.M. Meletinskii, D.M. Segal), on musical semiotics (B.M. Gasparov) and on lyric poetry (Zholkovskii, Revzin, Ivanov). Borrowing a favourite technique from the School's analytic procedures, description by analysis of pairs of equivalent elements in a system, one may treat the group's concerns by focusing on the work of two leading members, one from each centre: Iu.M. Lotman, by training as a literary scholar and intellectual historian (he had studied in Leningrad with surviving formalists) and B.A. Uspenskii, by training as a linguist and Slavic philologist (he had studied briefly in Denmark). Lotman's primary research area was Russian literature and culture of the 1780s to 1830s; Uspenskii's was the Russian language of the late medieval period and of the eighteenth century, although each ranged far afield in search of telling instances and challenging sign systems. By the late 1960s, however, each had moved towards the interests of the other, towards the investigations of ideology and point of view that marked Lotman's early work, towards studies of multiple linguistic codes that had interested Uspenskii, and some of their best work was collaborative.

Lotman's changing interests may be seen in four studies of Pushkin's novel in verse, *Evgenii Onegin*. The first (1960) analysed the construction of character in the novel; this piece is richly contextual, relating character development to contemporary trends in Russian thought, but treating the novel as an 'organic whole'. In 1966 he took a radically different approach in a paper on the artistic structure of *Evgenii Onegin*: here the operative concepts are modelling, a hierarchy of relationships, multiple points of view, system and anti-system, relationships, binary oppositions, and the inevitability of multiple interpretations given the text's multiplication of segments through the organization of its complex stanzas. Interpretation and the reader's role in the literary process become prominent, although they are governed by the internal relationships of the text. Lotman published a separate monograph on the novel in 1975. Here he achieves a synthesis of his work in intellectual history, his

semiotic analysis of paired opposition, and his Bakhtin-inspired study of multiple points of view as embodied in multiple discourses. He continues to study relations between the literary text and what lies outside it and the text's potential impact on a reader. In 1980 Lotman's commentaries to the novel treated culture in an anthropological sense, as the rituals, customs and conventions of Pushkin's Russia. But analysis by binary opposition - gentry/folk, Russian/European - remains an operative principle.

Uspenskii's work follows a similar trajectory, moving from the analysis of point of view towards studies of Russian culture in a broad, anthropological sense. His *O semiotike ikony (Semiotics of the Russian Icon)* (1971) subjects the 'language' of icons to linguistic analysis (syntactic, semantic, pragmatic), paying particular attention to the 'internal perspective' of medieval icons, according to which painter and viewer establish their point of view within the painting, not outside it. *Poetika kompozitsii (A Poetics of Composition)* (1970) had treated point of view more generally, and, although terms are not always defined with desirable precision, the book suggests an innovatory alternative to plot-based narratology, namely, the analysis with respect to four levels of point of view (ideological, phraseological, spatial and temporal, psychological).

For Lotman, as for Uspenskii, the text is both a sign (in the secondary modelling system) and a sequence of signs in verbal language. Both continued to focus their attention on problems of communication and point of view. Lotman's *Struktura khudozhestvennogo teksta (Structure of the Artistic Text)* (1970) and *Analyz poeticheskogo teksta (Analysis of the Poetic Text)* (1972) share some of the same problems of definition and focus that appeared in Uspenskii's *A Poetics of Composition*. But Lotman, the literary scholar, pays more attention to linguistic analysis and to information theory (the latter generally used metaphorically) than the linguist Uspenskii. While problems of integration in moving from level to level and in moving from analysis to interpretation are not solved theoretically, these books offer many suggestive ideas and fragments of literary analysis. Lotman's treatment of 'event' in the first book as 'the shifting of a character across the borders of a semantic field', as a transgression, or as the violation of an expectation, has been widely influential in studies of narrative.

From these large synthetic works on artistic texts, Lotman and Uspenskii moved to speculative discussions on an even higher level of synthesis, the typology of culture. In their 1971 essay, 'On the Semiotic Mechanism of Culture', culture becomes a system of sign production marked off from non-culture, a system of constraints and prescriptions. Here they differentiate cultures by modes of sign production and, predictably, Lotman and Uspenskii identify two opposed types of culture: cultures oriented towards expression (which view themselves in terms of an aggregate of texts) and cultures oriented towards content (which see themselves as a system of rules). In a more specific and controversial essay, 'Binary Models in the Dynamics of Russian Culture' (1977), they treat Russian culture as marked by a binary opposition between the 'sacred' and 'profane', an opposition which dictates diametrically opposed modes of behaviour. The same terms, for instance new/old, can in different periods occupy different positions in this hierarchy. Lotman and Uspenskii oppose this binary structure to a European ternary one, in which a neutral middle sphere allows for new systems to develop gradually, not catastrophically.

Subsequent research by Lotman and Uspenskii in the concrete phenomena in Russian cultural history address topics of moral and political significance: lying, theatricality, obscenity, religious dissent, pretenders to the throne and the use of multiple languages. Lotman's penultimate book, *Kul'tura i vzryv (Culture and Explosion)* (1992), published just after the fall of the Soviet Union, returns to speculative semiotic-based historiography to re-examine the relationship of culture to what is outside it and to explore Russia's new possibility of entering the 'European ternary system', with its concomitant development by gradual change as opposed to change by violent overthrow. Lotman's last book, his *Besedy o russkoi kul'ture (Conversations About Russian Culture)* (1994), returns to the turn of the nineteenth century to give a thick description of the daily life and traditions of the Russian gentry - its rituals, fashions and social patterns and the meaning that these had for those who practised them. The principled theoretical argumentation of these books, and their rigorous wide-ranging research and openness to experiment, encompass the main achievement of the Moscow-Tartu School.

See also: Semiotics; Structuralism in linguistics

WILLIAM MILLS TODD III

List of works

Baran, H. (ed.) (1976) *Semiotics and Structuralism: Readings from the Soviet Union*, White Plains: IASP.

(Well-introduced selection of papers on culture, mythology and literature.)

Lotman, Iu.M. (1960) 'K èvoliutsii postroeniia kharakterov v romane ''Evgenii Onegin''' (On the Evolution of the Construction of Characters in the Novel *Eugene Onegin*), in *Pushkin: Issledovaniia i materialy*, vol. 3, Moscow-Leningrad: Nauka, 131-73.(An excellent example of Lotman's early work in intellectual history.)

Lotman, Iu.M. (1964) *Lektsii po struktural'noi poëtike (Lectures on Structural Poetics)*, in *Trudy po znakovym sistemam (Works on Sign Systems)*, vol. 1, Tartu: Uchënye zapiski Tartuskogo Gos. Universiteta, No. 160. (Introduction to the structure of verse.)

Lotman, Iu.M. (1966) 'Khudozhestvennaia struktura ''Evgenii Onegin''' (The Artistic Structure of 'Eugene Onegin') in *Uchënye zapiski Tartuskogo Gos. Universiteta* 184: 5-32.(An important early example of Lotman's structuralist analysis.)

Lotman, Iu.M. (1970) *Struktura khudozhestvennogo teksta*, Moscow; trans. G. Lenhoff and R. Vroon, *The Structure of the Artistic Text*, Ann Arbor, MI: Michigan Slavic Contributions, 1977.(The most complete discussion of structuralist poetics, together with topics relevant to prose, such as plot and character; difficult to read.)

Lotman, Iu.M. (1972) *Analyz poeticheskogo teksta*, Leningrad; trans. D.B. Johnson, *Analysis of the Poetic Text*, Ann Arbor, MI: Ardis, 1976.(Further development of the theory of verse together with close analyses of a dozen lyric poems.)

Lotman, Iu.M. (1973) *Semiotika kino i problemy kinoëstetiki*, trans. M.E. Suino, *Semiotics of Cinema*, Ann Arbor, MI: Michigan Slavic Contributions, 1976.(A brief, accessible outline of selected topics in a grammar of cinematography.)

Lotman, Iu.M. (1975) *Roman v stikhakh Pushkina 'Evgenii Onegin' (Pushkin's Novel in Verse 'Eugene Onegin')*, Tartu.(An example of Lotman's structuralist analysis, more extensive and less technical than the 1966 article on *Eugene Onegin*.)

Lotman, Iu.M. (1980) *Roman A.S. Pushkina 'Evgenii Onegin': Kommentarii (A.S. Pushkin's Novel 'Eugene Onegin': Commentaries)*, Leningrad.(Semiotic analysis, tending towards cultural anthropology; very accessible.)

Lotman, Iu.M. (1990) *Universe of the Mind: A Semiotic Theory of Culture*, trans. A. Shukman, Bloomington, IN: Indiana University Press.(A well-integrated collection of papers on text, rhetoric, space and culture.)

Lotman, Iu.M. (1992) *Kul'tura i vzryv (Culture and Explosion)*, Moscow: Gnosis.(A semiotic approach to culture and historical transformation.)

Lotman, Iu.M. (1992-3) *Izbrannye stat'i (Selected Articles)*, Tallinn: Aleksandra, 3 vols.(Prepared by Lotman himself, includes essays on semiotics, cultural history and literature.)

Lotman, Iu.M. (1994) *Besedy o russkoi kul'ture: Byt i traditsii russkogo dvorianstva (XVIII-nachalo XIX veka) (Conversations About Russian Culture: Everday life and traditions of the Russian gentry (18th to early 19th century))*, St Petersburg: Iskusstvo.(Popular, easily accessible discussions of such topics as duels, cardplaying, balls and gender relations.)

Lucid, D.P. (1977) *Soviet Semiotics: An Anthology*, Baltimore, MD: Johns Hopkins University Press.(Papers on general concepts, modelling systems, communication, art and culture.)

Matejka, L. *et al.*, (eds) (1977) *Readings in Soviet Semiotics (Russian Texts)*, Ann Arbor, MI: Michigan Slavic Publications.(Excellent selection, especially of the School's early papers.)

Nakhimovsky, A.D. and A.S. (eds) (1985) *The Semiotics of Russian Cultural History: Essays by Iurii M. Lotman, Lidiia Ia. Ginzburg, Boris A. Uspenskii*, Ithaca, NY: Cornell University Press.(Well-chosen papers on cultural typology and history.)

Paperno, I. (1988) *Chernyshevsky and the Age of Realism: A Study in the Semiotics of Behavior*, Stanford, CA: Stanford University Press.(Highly accessible and intelligent development of the School's ideas on text, behaviour, modelling and culture.)

Shukman, A. (ed.) (1978) *Soviet Semiotics and Criticism: An Anthology*, special issue of *New Literary History* 9 (3).(Includes 'On the Semiotic Mechanism of Culture' and well-chosen studies in the semiotics of literature.)

Shukman, A. (ed.) (1984) *The Semiotics of Russian Culture*, Ann Arbor, MI: Michigan Slavic Contributions. (Comprehensive set of papers on topics in cultural history.)

Tsiv'ian, T.V. (1987) *Issledovaniia po strukture teksta (Studies in the Structure of the Text)*, Moscow: Nauka. (Important papers by V.V. Ivanov and others on the theory, linguistics and analysis of texts.)

Uspenskii, B.A. (1970) *Poetika kompozitsii*, Moscow; trans. V. Zavarin and S. Wittig, *A Poetics of Composition*, Berkeley, CA: University of California, 1973.(An original approach to artworks in terms of point of view.)

Uspenskii, B.A. (1971) *O semiotike ikony (The Semiotics of the Russian Icon)*, Tartu; ed. S. Rudy, Lisse: Peter de Ridder, 1976.(An important application of semiotics to a non-verbal art form.)

Uspenskii, B.A. (1994a) *Izbrannye trudy (Selected Works)*, Moscow: Gnosis, 2 vols.(Selected papers on semiotics and cultural history.)

Uspenskii, B.A. (1994b) *Kratkii ocherk istorii russkogo literaturnogo iazyka (XI-XIX vv.) (A Brief Outline of the History of the Russian Literary Language (eleventh-nineteenth centuries))*, Moscow: Gnosis.(Important, controversial papers on diglossia and bilingualism.)

References and further reading

Gasparov, B. (1985) 'Introduction', in A.D. and A.S. Nakhimovsky (eds) *The Semiotics of Russian Cultural History: Essays by Iurii M. Lotman, Lidiia Ia. Ginzburg, Boris A. Uspenskii*, Ithaca, NY: Cornell University Press.(An elegant, insightful account of the School's history and work on culture.)

Ivanov, V.V. (1976) *Ocherki po istorii semiotiki v SSSR (Outlines of the History of Semiotics in the USSR)*, Moscow: Nauka.(An important attempt to draw a genealogy for the School.)

Koshelev, A.D. (1994) *Iu.M. Lotman i Tartusko-moskovskaia semioticheskaia shkola (Iu.M. Lotman and the Tartu-Moscow School of Semiotics)*, Moscow: Gnosis.(Includes memoirs by members of the School, and a bibliography of Tartu works on semiotics, 1964-92.)

Seyffert, P. (1985) *Soviet Literary Structuralism. Background. Debate. Issues*, Columbus, OH: Slavica.(An invaluable, exhaustive discussion of the Soviet debates about semiotics up to 1971.)

Shukman, A. (1977) *Literature and Semiotics: A Study of the Writings of Yu.M. Lotman*, Amsterdam: North Holland. (The most comprehensive and intelligent survey of Lotman's work, together with useful lists of papers from the Gorky conference, the 1962 symposium, and the first four summer schools.)

Performatives

There are certain things one can do just by saying what one is doing. This is possible if one uses a verb that names the very sort of act one is performing. Thus one can thank someone by saying 'Thank you', fire someone by saying 'You're fired' and apologize by saying 'I apologize'. These are examples of 'explicit performative utterances', statements in form but not in fact. Or so thought their discoverer, J.L. Austin, who contrasted them with 'constatives'. Their distinctive self-referential character might suggest that their force requires special explanation, but it is arguable that performativity can be explained by the general theory of speech acts.

In *How to Do Things with Words* (1962), J.L. Austin challenges the common philosophical assumption that indicative sentences are necessarily devices for making statements. Just as nonindicative sentences are marked grammatically for nonassertive use (interrogative sentences for asking questions and imperative sentences for requesting or ordering), certain *indicative* sentences are also so marked. Austin's paradigm is any sentence beginning with 'I' followed by an illocutionary verb (see Speech acts §1), such as 'promise', 'apologize' or 'request', in the simple present tense and active voice. One can make a promise by uttering the words 'I promise to go', but not by uttering 'I promised to go' or 'She promises to go'. The first-person plural can be performative too, as in 'We apologize… ', and so can the second-person passive, as in 'You're fired'. The word 'hereby' may be used before the performative verb to indicate that the utterance in which it occurs is the vehicle of the performance of the act in question.

Austin contended that these 'explicit performative utterances' are, unlike 'constatives' (statements, predictions, hypotheses, and so on), neither true nor false. In saying 'I promise to go' one is making a promise, not stating that one is making it. A performative promise is not, and does not involve, the statement that one is promising. It is an act of a distinctive sort, the very sort (promising) named by the performative verb. And, according to Austin, making explicit what one is doing is not describing what one is doing or stating that one is doing it.

Now it is also possible to promise without doing so explicitly, without using the performative verb, and this raises the question of whether there is a theoretically important difference between promising explicitly and doing it inexplicitly. A superficial difference is that in uttering the words 'I promise to go' the speaker is saying that he is promising to go and that this, what he is saying, is assessable as true or false. It is true just in case he is doing what he says he is doing - that is, promising to go. In general, in making an explicit performative utterance the speaker is saying what he is doing - and is thereby doing it. Does this mean that performativity requires a theoretically special explanation?

One suggestion is that performativity is a matter of linguistic meaning. Perhaps there is a special semantic property of performativity, so that it is part of the meaning of words like 'promise', 'apologize' and 'request' that one can perform an act of that very sort by uttering a performative sentence containing that verb. One problem with this suggestion is that it implausibly entails that such verbs are systematically ambiguous. For a performative sentence can be used literally but nonperformatively - for example, to report some habitual act. For instance, one might say 'I apologize… ' to describe typical situations in which one apologizes. Moreover, it seems that even if verbs like 'promise', 'apologize' and 'request' were never used performatively, they would still mean just what they mean in fact. Imagine a community of users of a language just like English in which there is no practice of using such verbs performatively. When people there perform acts of the relevant sorts, they always do so, just as we sometimes do, without using performative verbs, such as making promises by saying 'I will definitely… ', giving apologies by saying 'I'm sorry', and issuing requests by using imperative sentences. In this hypothetical community the verbs 'promise', 'apologize' and 'request' would seem to have the same meanings that they in fact have in English, applying, respectively, to acts of promising, apologizing and requesting. The only relevant difference would be that such acts are not performed by means of the performative form. It seems, then, that in our community, where they are sometimes performed in this way, performativity is not a matter of meaning.

It might be suggested that a special sort of convention is required for uses of the performative form to count as promises, apologies, requests, and so on. Explaining this by appealing to convention is gratuitous, however, for performativity is but a special case of a more general phenomenon. There are all sorts of other forms of words which are standardly used to perform speech acts of types not predictable from their semantic content, such as 'It

would be nice if you… ' to request, 'Why don't you… ?' to advise, 'Do you know… ?' to ask for information, 'I'm sorry' to apologize, and 'I wouldn't do that' to warn. In particular, there are hedged and embedded performatives, such as 'I can assure you… ', 'I must inform you… ', 'I would like to invite you… ', and 'I am pleased to be able to offer you… ', utterances to which the alleged conventions for simple performative forms could not apply. Could such conventions be suitably generalized? The variety of linguistic forms standardly used for the indirect performance of such speech acts seems too open-ended to be explained by any convention (or set of conventions) that is supposed to specify just those linguistic forms whose utterance counts as the performance of an act of the relevant sort.

An alternative explanation is needed. In general, speech acts are acts of communication, whose success in that regard requires the audience to identify the speaker's intention (see Speech acts §2). As Bach and Harnish (1979) have argued, what is special about the use of standardized forms of words, such as those illustrated above, is not that they provide the precedent for any conventions. Rather, the precedent provided by standardization serves to streamline the inference required on the part of the audience.

Convention seems relevant to performativity only in certain institutional contexts, where a specific form of words is designated, and often required, for the performance of an act of a certain sort. This is true of those performative utterances involved in, for example, adjourning a meeting, sentencing a criminal, or christening a ship. However, ordinary performative utterances are not bound to particular institutional situations. Like most speech acts, they are acts of communication and succeed not by conformity to convention but by recognition of intention (see Communication and intention).

See also: Austin, J.L.; Searle, J.R.; Speech acts

<div align="right">KENT BACH</div>

References and further reading

Austin, J.L. (1962) *How to Do Things with Words*, Cambridge, MA: Harvard University Press.(Formulates the distinction between performative and constative utterances and proposes a convention-based account of their successful felicitous performance.)

Bach, K. and Harnish, R.M. (1979) *Linguistic Communication and Speech Acts*, Cambridge, MA: MIT Press. (Chapter 10 argues that performativity can be explained inferentially, without appealing to any special meanings or conventions connected with performative verbs.)

Bach, K. and Harnish, R.M. (1992) 'How Performatives Really Work', *Linguistics and Philosophy* 15: 93-110. (Critically examines Searle's conventionalist approach and defends the inferentialist account.)

Recanati, F. (1987) *Meaning and Force: The Pragmatics of Performative Utterances*, Cambridge: Cambridge University Press.(Surveys the literature on performatives and presents an inferentialist account of how they work.)

Searle, J.R. (1989) 'How Performatives Work', *Linguistics and Philosophy* 12: 535-58.(Critically examines the inferentialist approach and defends a conventionalist account.)

Post-structuralism

Post-structuralism is a late-twentieth-century development in philosophy and literary theory, particularly associated with the work of Jacques Derrida and his followers. It originated as a reaction against structuralism, which first emerged in Ferdinand de Saussure's work on linguistics. By the 1950s structuralism had been adapted in anthropology (Lévi-Strauss), psychoanalysis (Lacan) and literary theory (Barthes), and there were hopes that it could provide the framework for rigorous accounts in all areas of the human sciences.

Although structuralism was never formulated as a philosophical theory in its own right, its implicit theoretical basis was a kind of Cartesianism, but without the emphasis on subjectivity. It aimed, like Descartes, at a logically rigorous system of knowledge based on sharp explicit definitions of fundamental concepts. The difference was that, for structuralism, the system itself was absolute, with no grounding in subjectivity. Post-structuralist critiques of structuralism typically challenge the assumption that systems are self-sufficient structures and question the possibility of the precise definitions on which systems of knowledge must be based.

Derrida carries out his critique of structuralist systems by the technique of deconstruction. This is the process of showing, through close textual and conceptual analysis, how definitions of fundamental concepts (for example, presence versus absence, true versus false) are undermined by the very effort to formulate and employ them. Derrida's approach has particularly influenced literary theory and criticism in the USA. In addition, Richard Rorty, developing themes from pragmatism and recent analytic philosophy, has put forward a distinctively American version of post-structuralism.

1 Structuralism

In his lectures on linguistics, Ferdinand de Saussure proposed a view of language (*langue*) as a formal structure, defined by differences between systemic elements. According to Saussure, this structure is simultaneously present in and unites the two domains of thought and words. A given linguistic term (a sign) is the union of an idea or concept (the 'signified') and a physical word (the 'signifier'). A language is a complete system of such signs, which exists not as a separate substance but merely as the differentiating form that defines the specific structure of both signifiers (physical words) and signifieds (ideas). Saussure's view rejects the common-sense picture of the set of signifiers and the set of signifieds as independent givens, with the signifieds having meaning in their own right and the signifiers obtaining meaning entirely through their association with corresponding signifieds. Saussure denies this independence and instead maintains that signifiers and signifieds alike have meaning only in virtue of the formal structure (itself defined by differences between elements) that they share (see Structuralism in linguistics).

Saussure's structuralist approach was very successful within linguistics, where it was applied and extended by, among others, Jakobson and Troubetzkoy. By the 1950s the approach had been adapted in anthropology (Lévi-Strauss), psychoanalysis (Lacan), and literary theory (Barthes); and there were hopes that it could provide the framework for rigorous accounts in all areas of the human sciences. Three distinguishing features of this framework were: (1) a rejection of all idealist views of concepts and meanings as derived from the activity of consciousness; (2) an understanding of concepts and meanings as, instead, grounded in the structural relations among the elements of abstract systems; (3) an explication of such structural relations solely in terms of bipolar differences (for example, real/unreal, temporal/nontemporal, present/absent, male/female).

2 Post-structuralism: terminology

Post-structuralism is obviously closely tied to structuralism, but commentators have characterized the relationship in a variety of mutually inconsistent ways. Some writers make no distinction between structuralism and post-structuralism, applying the single term 'structuralist' to the entire range of thinkers from Saussure through to Derrida. More commonly, post-structuralism is distinguished as a separate development, but there is disagreement as to whether it is primarily a reaction against structuralism or an extension of it (as the term *Neostrukturalismus*, commonly used by Manfred Frank and other German commentators, suggests). Apart from matters of definition, there is even disagreement as to whether major figures such as Barthes, Lacan and Foucault are structuralists or post-structuralists.

Michel Foucault's book, *Les mots et les choses* (translated under the title *The Order of Things*) is an instructive example. In one sense it is quintessentially structuralist. The book first uncovers the fundamental epistemic systems (which Foucault calls 'epistemes') that underlie and delimit the subjective thought of particular eras. It then goes on to show how the apparent ultimacy of subjectivity is itself just the product of one contingent episteme, that of modernity, which is even now disappearing (the famous 'death of man'). Nevertheless, Foucault's essentially historical viewpoint in the work demonstrates the limitation of structuralism: its inability to give any account of the transitions from one system of thought to another. Foucault seems to have seen from the beginning that structuralism can not be historical, a fact that explains his constant insistence that he was not a structuralist, in spite of his obvious deployment of structuralist methods and concepts. So, although *Les mots et les choses* is a structuralist book, it at the same time makes clear the limits of structuralism and prepares the way for Foucault's later work on power and ethics which is distinctly post-structuralist (see Foucault, M.; Post-structuralism and the social sciences).

Despite these ambiguities and disagreements, the concept of post-structuralism is useful, if not essential, for understanding philosophy in France during the latter part of the twentieth century. One fruitful approach is to think of post-structuralism as a philosophical reaction to the structuralism that was such a powerful force during the 1960s in linguistics, psychology and the social sciences. It was neither a simple rejection or extension of structuralism but a series of philosophical reflections on the structuralist programme and achievement.

3 Two major post-structuralist theses

Although structuralism was never formulated as a philosophical theory in its own right, its implicit theoretical basis was, as noted above, a kind of Cartesianism without the subject. (Hence, the association of structuralism with the notion of the 'death of the subject'.) Post-structuralist critiques of structuralism are typically based on two fundamental theses: (1) that no system can be autonomous (self-sufficient) in the way that structuralism requires; and (2) that the defining dichotomies on which structuralist systems are based express distinctions that do not hold up under careful scrutiny.

The first thesis is not understood so as to support the traditional idealist view that systematic structures are dependent on the constitutive activities of subjects. Post-structuralists retain structuralism's elimination of the subject from any role as a foundation of reality or of our knowledge of it. But, in opposition to structuralism, they also reject any logical foundation for a system of thought (in, for example, its internal coherence). For post-structuralists, there is no foundation of any sort that can guarantee the validity or stability of any system of thought.

The second thesis is the key to post-structuralism's denial of the internal coherence of systems. The logical structure of a system requires that the applications of its concepts be unambiguously defined. (In the formalism of elementary number theory, for example, there must be no question as to whether a given number is odd or even.) As a result, the possibility of a systematic structure depends on the possibility of drawing sharp distinctions between complementary concepts such as odd/even, charged/uncharged, living/non-living, male/female and so on. Post-structuralist philosophers have been particularly concerned with the fundamental dichotomies (or oppositions) underlying structuralist theories in the human sciences. Saussure's linguistics, for example, is based on the distinction of the signifier from the signified; Lévi-Strauss's anthropology of myths employs oppositions such as raw/cooked, sun/moon and so on. In each case, post-structuralists have argued that the dichotomy has no absolute status because the alternatives it offers are neither exclusive nor exhaustive.

4 Derrida's critique of logocentrism

This sort of critique was extended to philosophy, particularly by Jacques Derrida, who finds Western philosophical thought pervaded by a network of oppositions - appearance/reality, false/true, opinion/knowledge, to cite just a few examples - that constitute what he calls the system of 'logocentrism'. This term derives from Derrida's conviction that at the root of Western philosophical thought is a fundamental distinction between speech (*logos*) and writing. Speech is privileged as the expression of what is immediate and present, the source, accordingly, of what is real, true and certain. Writing, on the other hand, is derogated as an inferior imitation of speech, the residue of speech that is no longer present and, therefore, the locus of appearance, deceptions and uncertainty. Plato's devaluation of writing in comparison with living dialogue is the most famous and influential example of this distinction. But

Derrida finds the distinction pervading Western philosophy and regards it as not just a preference for one form of communication over another but the basis for the entire set of hierarchical oppositions that characterize philosophical thought. Speech offers presence, truth, reality, whereas writing, a derivative presentation employed in the absence of living speech, inevitably misleads us into accepting illusions.

Derrida's critiques of the speech/writing opposition - and of all the hierarchical oppositions that attend it - proceed by what he calls the method of 'deconstruction' (see Deconstruction). This is the process of showing, through close textual and conceptual analysis, how such oppositions are contradicted by the very effort to formulate and employ them. Consider, for example, the opposition between presence and absence, which plays a fundamental role in Husserl's phenomenology (and many other philosophical contexts). Husserl requires a sharp distinction between what is immediately present to consciousness (and therefore entirely certain) and what is outside of consciousness (and therefore uncertain). But once Husserl undertakes a close analysis of the immediately present, he discovers that it is not instantaneous but includes its own temporal extension. The 'present', as a concrete experiential unit, involves both memory of the just-immediately-past (retention, in Husserl's terminology) and anticipation of the immediate future (protention). Thus, the past and the future, both paradigms of what is absent (not present), turn out to be integral parts of the present. Husserl's own account of the presence/absence opposition overturns it.

Deconstruction maintains that there is no stability in any of thought's fundamental oppositions. Their allegedly exclusive alternatives turn out to be inextricably connected; their implicit hierarchies perpetually reversible. As a result, there is an ineliminable gap between the intelligibility of a rational system and the reality it is trying to capture. Derrida expresses this gap through a variety of terms. He frequently speaks of *différance* (a deliberately misspelled homophone of the French *différence*) to emphasize, first, the difference between systematic structures and the objects (for example, experiences, events, texts) they try to make intelligible, and, second, the way in which efforts to make absolute distinctions are always deferred (another sense of the French *différer*) by the involvement of one polar opposite in the other. This latter phenomenon Derrida also discusses in terms of the 'trace' of its opposite always lingering at the heart of any polar term. He also employs the term 'dissemination' to refer to the way that objects of analysis slip through the conceptual net spread by any given system of intelligibility we devise for it.

5 Post-structuralism and literary theory

Thus far the discussion has focused on Derrida's deconstruction of the meaningful structures philosophers purport to find in reality and to express in their philosophical texts. But Derrida's approach is also readily applicable to literary texts (and the 'worlds' they create). This is because - like philosophical systems - poems, novels and other literary texts are typically thought to embody complete and coherent systems of meaning, which it is the task of literary criticism to extract. Although Derrida himself has dealt primarily with philosophical texts, his approach has been widely adopted by analysts of literature. (Of course, as should be expected, Derrida and his followers reject any sharp distinction between the philosophical and the literary.)

Traditional literary analysis has understood the meaning of a text as the expression of its author's mind; that is, as thoughts the author intended to convey in writing the text. The first stage of deconstructive criticism is the structuralist one of detaching meaning from authorial intention, locating it instead in the text itself as a linguistic structure. Roland Barthes, for example, showed how to analyse a text by Balzac entirely in terms of the formal codes it embodies, with no reference to what Balzac supposedly 'meant'. This structuralist move effects a 'death of the author' parallel to the anti-Cartesian 'death of the subject'. But the post-structuralists take the future step of denying a fixed meaning to even the autonomous text itself. It is not that a text lacks all meaning but that, on the contrary, it is the source of an endless proliferation of conflicting meanings. As deconstructionists delight in showing, any proposed privileged meaning of a text can be undermined by careful attention to the role in it of apparently marginal features. (For example, an orthodox Christian reading of Milton's *Paradise Lost* is deconstructed by a close study of certain details in its treatment of Satan.) There is no doubt, of course, that texts are often produced by authors trying to express what they think or feel. But what they write always goes beyond any authorial intention and in ways that can never be reduced to a coherent system of meaning.

The deconstructionist's point can also be understood as an undermining of the distinction between primary text and commentary. On the traditional view, a commentary is an effort to formulate as accurately as possible the

content (meaning) of the text. To the extent that it is successful, a commentary expresses nothing more and nothing less than this meaning. But for deconstructionists the meaning in question does not exist, and the commentary must be understood as nothing more than a free elaboration of themes suggested, but not required, by the text. Unable to be a secondary reflection, the commentary becomes as much an independent creation as the text itself.

6 Rorty's post-structuralist pragmatism

Richard Rorty's work is far removed, in both antecedents and style, from that of continental post-structuralists (see Rorty, R.). His critique of Cartesianism, derived more from Dewey than from Heidegger, is aimed at twentieth-century analytic philosophy rather than the structuralist human sciences; and his urbanely lucid prose contrasts sharply with the wilfully playful convolutions of Derrida and his followers. None the less, Rorty's analyses lead him to a critique of traditional philosophy very similar to that of the post-structuralists.

The focal point of Rorty's critique is the project (called foundationalism) of providing a philosophical grounding for all knowledge. Modern foundationalism originates with Descartes, but Rorty sees it as also the *leitmotif* of Descartes' successors, through Hume and Kant down to the logical positivists. Like Derrida, Rorty attacks traditional systematic thought by calling into question some of its key distinctions. Unlike Derrida, however, he does not carry out his attacks through close readings of classic texts but by deploying the results of recent analytic philosophy. He uses, for example, Quine's critique of the distinction between analytic and synthetic statements to argue that there are no foundational truths about the meaning of concepts. He appeals to Wilfrid Sellars's undermining of the distinction between theory and observation to reject empirical foundations of knowledge in interpretation-free sense data. He employs Donald Davidson's questioning of the distinction between the formal structure and the material content of a conceptual framework to reject Kantian attempts to ground knowledge in principles that define the framework of all possible thought.

In Rorty's view the upshot of these various critiques is to cut off every source of an ultimate philosophical foundation for our knowledge. Accordingly, he maintains, philosophy must give up its traditional claim to be the final court of appeal in disputes about truth. We have no alternative but to accept as true what we (the community of knowers) agree on. There is no appeal beyond the results of the 'conversation of mankind' so far as it has advanced to date. For us, there is no (upper-case) Truth justified by privileged insights and methods. There is only the mundane (lower-case) truth: what our interlocutors let us get away with saying.

It might seem that this rejection of foundationalism is a rejection of the entire tradition of Western philosophy since Plato. Rorty, however, distinguishes two styles of philosophy. First, there is *systematic philosophy*, the mainline of the Western tradition since Plato, which is defined by the foundationalist goal of ultimate justification. But, on the other hand, there is another enterprise, always marginal to the tradition, that Rorty calls *edifying philosophy*. Whereas systematic philosophers undertake elaborate and purportedly eternal constructions (which are always demolished by the next generation), edifying philosophers are content to shoot ironic barbs at the systematic thought of their day, exploding its pretensions and stimulating intriguing lines of counter-thought. The tradition of edifying thought can be traced back at least to the ancient Cynics and has been more recently represented by Kierkegaard, Nietzsche and the later Wittgenstein. Derrida's deconstructions are, on Rorty's view, a prime contemporary example of edifying philosophy.

Edifying philosophers, however, are philosophers only because they react against systematic philosophy. They do not differ from other sorts of cultural critics (novelists, literary theorists, social scientists) because of any distinctively philosophical method or viewpoint. If, in the wake of thinkers such as Derrida and Rorty, systematic philosophy is abandoned, philosophy will be too. The triumph of post-structuralism would, for better or worse, be the end of philosophy as we have known it.

See also: Postmodernism

<div align="right">GARY GUTTING</div>

References and further reading

Culler, J. (1982) *On Deconstruction: Theory and Criticism after Structuralism*, Ithaca, NY: Cornell University Press.(Excellent discussion of Derrida's views and of their impact on literary criticism.)

Dews, P. (1987) *Logics of Disintegration: Poststructuralist Thought and the Claims of Critical Theory*, London:

Verso.(Analytic and critical essays on Derrida, Lacan, Foucault and Lyotard.)

Foucault, M. (1966) *Les mots et les choses: une archéologie des sciences humaines*, Paris: Gallimard; trans. A. Sheridan, *The Order of Things*, New York: Random House, 1970.(A critical history of the origins of the modern social sciences.)

Frank, M. (1989) *What Is Neostructuralism?*, trans. S. Wilke and R. Gray, Minneapolis, MN: University of Minnesota Press.(A detailed study and critique of the movement.)

Rorty, R. (1989) *Contingency, Irony, and Solidarity*, Cambridge: Cambridge University Press.(A collection of essays that formulates Rorty's position and comments on the views of other post-structuralists.)

Rorty, R. (1991) *Essays on Heidegger and Others*, Cambridge: Cambridge University Press.(Another collection of essays that formulates the author's position and comments on the views of other post-structuralists.)

Sturrock, J. (ed.) (1979) *Structuralism and Since*, Oxford: Oxford University Press.(A collection of introductory essays on the major figures.)

Pragmatics

Analytic philosophers have made lasting contributions to the scientific study of language. Semantics (the study of meaning) and pragmatics (the study of language in use) are two important areas of linguistic research which owe their shape to the groundwork done by philosophers.

Although the two disciplines are now conceived of as complementary, the philosophical movements out of which they grew were very much in competition. In the middle of the twentieth century, there were two opposing 'camps' within the analytic philosophy of language. The first - 'ideal language philosophy', as it was then called - was that of the pioneers, Frege, Russell and the logical positivists. They were, first and foremost, logicians studying formal languages and, through these formal languages, 'language' in general. Work in this tradition (especially that of Frege, Russell, Carnap, Tarski and later Montague) gave rise to contemporary formal semantics*, a very active discipline developed jointly by logicians, philosophers and grammarians. The other camp was that of so-called 'ordinary language philosophers', who thought important features of natural language were not revealed, but hidden, by the logical approach initiated by Frege and Russell. They advocated a more descriptive approach, and emphasized the 'pragmatic' nature of natural language as opposed to, for example, the 'language' of Principia Mathematica. Their own work (especially that of Austin, Strawson, Grice and the later Wittgenstein) gave rise to contemporary pragmatics, a discipline which (like formal semantics) has developed successfully within linguistics in the past thirty years.*

From the general conception put forward by ordinary language philosophers, four areas or topics of research emerged, which jointly constitute the core of pragmatics: speech acts; indexicality and context-sensitivity; non-truth-conditional aspects of meaning; and contextual implications. In the first half of this entry, we look at these topics from the point of view of ordinary language philosophy; the second half presents the contemporary picture. From the first point of view, pragmatics is seen as an alternative *to the truth-conditional approach to meaning associated with ideal language philosophy (and successfully pursued within formal semantics). From the second point of view, pragmatics merely* supplements *that approach.*

1 Pragmatics and ordinary language philosophy

The linguistic investigations undertaken by ordinary language philosophers in what was to become 'pragmatics' had been notably anticipated by various researchers belonging to other traditions (phenomenologists like Marty or Reinach, linguists like Bally or Gardiner, psychologists like Bühler, or anthropologists like Malinowski). However, what influenced ordinary language philosophers most was the conception of language advocated by 'ideal language philosophers', against which they reacted strongly (see Ordinary language philosophy).

Central in the ideal language tradition had been the equation of, or at least the close connection between, the meaning of a sentence and its truth-conditions. This truth-conditional approach to meaning, perpetuated by contemporary formal semantics, is one of the things which ordinary language philosophers found quite unpalatable (see Meaning and truth). Their own emphasis was on the distinction between 'language' and 'speech' (Gardiner 1932) or, equivalently, between 'sentence' and 'statement' (Austin [1950] 1971; Strawson [1950] 1971). It is the sentence (a unit of 'language') which has meaning, according to ordinary language philosophers; whereas it is the statement made by uttering the sentence in a particular context which has truth-conditions. The sentence itself does not have truth-conditions. Truth can only be predicated of sentences indirectly, via the connections between the sentence and the 'speech act' it can be used to perform. Rather than equating the meaning of a sentence with its alleged truth-conditions, some philosophers in the pragmatic tradition have suggested equating it with its speech act 'potential' (which may include, as a proper part, a certain truth-conditional potential; see Alston 1964: 37-9).

Suppose that we posit abstract objects, namely 'propositions', which have their truth-conditions essentially. Then the point made by ordinary language philosophers can be put as follows: sentences do not express propositions *in vacuo*, but only in the context of a speech act. Given that the same sentence can be used to make different speech acts with different contents, the 'proposition' which is the content of the speech act must be distinguished from the linguistic meaning of the sentence *qua* unit of the language ('sentence meaning') (see Propositions, sentences and statements). It must also be distinguished from the contextually determined meaning of a particular utterance of the sentence ('utterance meaning'), for the latter includes much more than merely the propositional content of the

speech act performed in uttering the sentence. Utterance meaning includes a rich 'non-truth-conditional' component: Besides the proposition it expresses, an utterance conveys indications concerning the type of speech act being performed, the attitudes of the speaker, the place of the utterance within the discourse, its presuppositions and so forth. Moreover there is a secondary layer of meaning which includes the 'contextual implications' of the speech act, and in particular what H.P. Grice called the 'conversational implicatures' of the utterance.

2 Speech acts

Speech act theory (see Austin 1975; Searle 1969) is concerned with communication: not communication in the narrow sense of transmission of information, but communication in a broader sense which includes the issuing of orders, the asking of questions, the making of apologies and promises, and so on. According to the theory, a speech act is more than merely the uttering of a grammatical sentence endowed with sense and reference. To speak is also to *do* something in a fairly strong sense: it is to perform what J.L. Austin called an 'illocutionary act'. In performing an illocutionary act, a speaker takes on a certain role and assigns a corresponding role to the hearer. By giving an order, speakers express the desire that their hearer follow a certain course of conduct and present themselves as having the requisite authority to oblige the hearer to follow the course of conduct in question, simply because it is their will. The social role taken on by the speaker who gives an order is embodied in the organizational notion of 'superior rank'. Austin stressed such institutional embodiments of illocutionary roles in order to show that language itself is a vast institution incorporating an array of conventional roles corresponding to the range of socially recognized illocutionary acts. From this point of view, assertion - the act of making a statement - is only one illocutionary act among many others.

Illocutionary acts have 'felicity conditions' (conditions which must be contextually satisfied for the illocutionary act to be successfully performed). Thus an assertion about an object 'presupposes' the existence of that object and is felicitous only if the object in question actually exists (Strawson [1950] 1971; Austin 1975). The study of felicity conditions is a central concern of speech act theory, along with the taxonomy of illocutionary acts. But the most central concern, perhaps, relates to the characterization of the very notion of an illocutionary act. Illocutionary acts are generally introduced ostensively, by examples, and they are distinguished both from the mere act of saying something ('locutionary act') and from the act of causing something to happen by saying something ('perlocutionary act', for example frightening, convincing and so on). The nature of the intermediate category of 'illocutionary acts' remains a matter of debate, however. The pioneers of speech act theory, Austin and Searle, advocated an institutional or conventional approach. In this framework the illocutionary acts performed in speech, like the acts that are performed in games (for example, 'winning a set' in tennis), are governed by rules and exist only against a background of conventions. But an alternative, 'intentionalist' view, originating from Grice [1957] 1989 and Strawson [1964] 1971, developed and is now the dominant trend in speech act theory (see §13) (see Speech acts).

3 Contextual implications

The notion of a contextual implication itself is a speech-act theoretic notion. If, besides the meaning or content of an utterance, there is another realm (namely that of the illocutionary act the utterance serves to perform), then along with the implications of what is said there will be a further set of implications derivable from the utterance (namely the implications of the illocutionary act itself). Some of these 'pragmatic' implications are fairly trivial. Thus, according to ordinary language philosophers, it is a rule of the language game of assertion that whoever asserts something believes what they say and has some evidence for it; even the liar, who does not obey this rule, has to pretend that they do if they want to participate in the game. This rule generates pragmatic implications: by asserting something and therefore engaging in the language game, the speaker 'implies' that they obey the rules of the game and, therefore, that they believe whatever they are asserting. The speaker cannot disavow these implications of their speech act without 'pragmatic contradiction'. A pragmatic contradiction is a conflict between what an utterance says and what it pragmatically implies. Thus Moore's famous paradoxical utterance, 'It is raining but I do not believe it', is not self-contradictory in the logical sense: the state of affairs it describes is logically possible (it might be raining without the speaker's knowing it). But the speaker's asserting that it rains implies that they believe it, and this contradicts the second part of the utterance. (The twin notions of pragmatic implication and pragmatic contradiction or 'pragmatic paradox' have been used to illuminate a variety of philosophical issues, including the nature of Descartes' *Cogito*.)

Less trivial are the contextual implications famously discussed by H.P. Grice (1989). According to Grice, the speaker making an utterance does not merely imply that they respect the rules of the language game; among the pragmatic implications of the utterance, we find a number of additional assumptions contextually required in order to maintain the supposition that the rules of the game are being observed. Suppose that I am asked whether I will go out; I reply: 'It is raining'. As stated above, it is a rule of assertion that the assertor believes what they say and have some evidence for it. By virtue of this rule, my utterance implies that I believe that it is raining, and that I have some evidence for my assertion. Considered as an answer to a question, my utterance also implies that it provides the information requested by the addressee, for it is a rule of the Question-and-Answer game that the answerer must provide the requested piece of information. Now in order to maintain the supposition that the speaker's utterance actually provides the requested information, additional premises are needed: for example, the assumption that the speaker will not go out if it rains. In conjunction with this contextual assumption, the utterance implies that the speaker will not go out, thereby providing a negative answer to the question. In so far as they serve to restore the utterance's conformity to the rules of the game, the conclusion that the speaker will not go out and the contextual assumption through which it is derived are further pragmatic implications of the utterance. Grice called them 'conversational implicatures'. Contrary to the more trivial pragmatic implications, they can be disavowed by the speaker without pragmatic contradiction (at least if there is another way of making the utterance compatible with the supposition that the rules of the game are being respected). This distinguishing feature of conversational implicatures is referred to as their 'cancellability'. Implicatures which are not disavowed are legitimately taken as part of what the utterance communicates. They constitute a second layer of meaning, additional to what is literally said (see §12) (see Implicature).

4 Non-truth-conditional aspects of meaning

Like pragmatic implications, non-truth-conditional aspects of meaning are easy to account for if speech is considered as a rule-governed activity (Stenius 1967). What is the meaning of, for example, the imperative mood? Arguably, the sentences 'You will go to the shop tomorrow at 8', 'Will you go to the shop tomorrow at 8?' and 'Go to the shop tomorrow at 8' describe the same (sort of) state of affairs. The difference between them is pragmatic rather than descriptive: it relates to the type of illocutionary act being performed by the utterance. Thus the imperative mood indicates that the speaker, in uttering the sentence, performs an illocutionary act of a 'directive' type. (Such an act is governed by the rule that if the speaker performs a directive act with content *P*, the addressee is to make it the case that *P*.) To account for this 'indication', which does not belong to the utterance's descriptive or propositional content, we can posit a rule or convention to the effect that the imperative mood is to be used only if one is performing a directive type of illocutionary act. This rule gives 'conditions of use' for the imperative mood. By virtue of this rule, a particular token of the imperative mood in an utterance u 'indicates' that a directive type of speech act is being performed by u. This (token-reflexive) indication conveyed by the token follows from the conditions of use that govern the type; these conditions of use constitute the linguistic meaning of the type (Recanati 1987: 15-7).

Pragmatic indications are a species of pragmatic implication: they are what the use of a particular expression pragmatically implies, by virtue of a certain condition of use conventionally associated with the expression. In contrast to more standard pragmatic implications, however, pragmatic indications are linguistically encoded, via the condition of use conventionally associated with the expression. Grice (1989) called such conventional pragmatic implications 'conventional implicatures', as opposed to 'conversational implicatures' (see Implicature §4). Whether they concern the type of the illocutionary act, as in the example I have given, or some other aspect of the context of utterance, pragmatic indications can always be accounted for in terms of conditions of use. They are 'use-conditional' aspects of meaning. Their exploration is one of the empirical tasks of semantics construed as the study of linguistic meaning under all its aspects (see §8 for an alternative construal of 'semantics').

5 Indexicals

Use-conditional meaning is not incompatible with descriptive content, in the sense that one and the same expression can be endowed with both. There are expressions which have a purely use-conditional meaning and do not contribute to truth-conditional content. Illocutionary markers like the imperative mood, or discourse particles such as 'well', 'still', 'after all', 'anyway', 'therefore', 'alas', 'oh' and so forth, fall into this category. Thus the following utterances have the same truth-conditional content, and are distinguished only by the pragmatic

indications they respectively convey:

Well, Peter did not show up
Still, Peter did not show up
After all, Peter did not show up
Therefore, Peter did not show up
Alas, Peter did not show up

But there are also expressions which have a two-layered meaning. Indexicals are a case in point. A 'rule of use' is clearly associated with indexicals: thus 'I' is governed by a convention of use (it is to be used to refer to the speaker). By virtue of this conventional rule, a use u of 'I' token-reflexively indicates that it refers to the speaker of u. But u also contributes to the utterance's truth-conditional content. 'I' being a directly referential expression, its truth-conditional contribution (its 'content') is its actual referent, not the rule of use which contextually determines the referent (Kaplan 1989: 481-563; Recanati 1993).

Besides the horizontal distinction between truth-conditional and non-truth-conditional aspects of meaning, we see that there is a vertical distinction between two levels of meaning for indexical expressions (Strawson [1950] 1971; Kaplan 1989: 481-563). At the first level - corresponding to the linguistic meaning of the expression-type - we find the rule of use conventionally associated with the expression. At the second level - corresponding to the context-dependent semantic value of the token - the rule of use determines the expression's 'content' (see Demonstratives and indexicals).

6 Levels of meaning

The two distinctions we have made, between truth-conditional and non-truth-conditional aspects of meaning on the one hand, and between levels of meaning on the other hand, should not be conflated (as they often have been). Despite appearances, they are orthogonal to each other. In the same way as the truth-conditional content of an indexical sentence is context-dependent and, therefore, belongs to the second level of meaning, the pragmatic indications conveyed by an expression governed by a rule of use also are context-dependent and belong to the second level of meaning. In other words, a distinction must be made between the rule of use (first level of meaning) and the pragmatic indications it contextually generates, in the same way as we distinguish between the rule of use and the truth-conditional content it contextually determines.

That pragmatic indications, though conventional, are context-dependent is shown by examples like (1):

(1) The weather is nice, but I have a lot of work.

The conjunction 'but' is governed by a certain condition of use which distinguishes it from 'and'. According to Ducrot (1972: 128-9), 'but' is to be used only if the following conditions are contextually satisfied:

(i) The first conjunct (P) supports a certain conclusion r;
(ii) The second conjuncts (Q) supports not-r;
(iii) Q is considered stronger than P, that is the whole utterance supports not-r.

Uttering (1) pragmatically implies that the conditions of use associated with 'but' are satisfied, that is that there is a conclusion r such that the first conjunct supports r and the second conjunct more strongly supports not-r. But the pragmatic implication conveyed by a particular utterance of (1) is much more specific. In context, the variable r is assigned a particular interpretation, for example 'we should go for a walk'. Example (1) therefore pragmatically implies something like the following: we should not go for a walk (because of all the work I have to do), despite the nice weather which suggests otherwise.

In so far as it is context-dependent and conveyed by the token, this pragmatic implication is to be located at the second level of meaning, alongside the content of indexicals. Even in a case where the pragmatic indication is fully conventional and not in need of contextual specification, it is conveyed by the token, not by the type. Thus a particular use u of the pronoun 'I' indicates that it (u) refers to the speaker of u. This token-reflexive indication is distinct from the rule of use, to the effect that for all x, if x is a token of 'I' it must be used to refer to the speaker of x.

The picture is further complicated by the Gricean distinction between what is literally said and what is non-literally or indirectly communicated. We end up with a three-fold distinction between the following layers of meaning:

First level. Rules of use ('character', in Kaplan's framework).
Second level. Truth-conditional content + pragmatic indications.
Third level. Conversational implicatures.

The need for a third level of meaning comes from the fact that the contextual process responsible for conversational implicatures (and non-conventional pragmatic implications in general) takes the second-level meaning of the utterance as input. When an expression is governed by a condition of use, using that expression pragmatically implies that the condition is satisfied. But conversational implicatures, in contrast to conventional implicatures, are not generated by virtue of a condition of use directly associated with a particular linguistic expression; they are normally generated by virtue of conversational norms that concern the content of utterances rather than the expressions which are used to convey that content. For example, a speaker should not say what he believes to be false (the 'maxim of quality' in Grice's terminology); as a result, saying that *P* pragmatically implies that the speaker believes that *P*. The generation of this pragmatic implication presupposes that the proposition expressed has been identified: from the fact that *the speaker has said that P*, together with the default assumption that the maxim of quality is respected, we can infer that *the speaker believes that P*. The implicature-generating process therefore deserves to be called a 'secondary pragmatic process' (Recanati 1993). There are three basic levels of meaning, with the context controlling the transition from the first to the second and from the second to the third. The proposition expressed by the utterance must first be contextually identified (primary pragmatic process) in order for the non-conventional pragmatic implications to be derived (secondary pragmatic process).

7 Open texture

For ordinary language philosophers, the truth-conditional or 'descriptive' content of an utterance is a property of the speech act, not a property of the sentence. A sentence only has truth-conditions in the context of a speech act. This is so not merely because of indexicality: the fact that the reference of some words depends on the context in a systematic way. Indexicality is only one form of context-dependence. There is another one, no less important, which affects the sense (the conditions of application) of words rather than their reference. According to Austin and Wittgenstein, words have clear conditions of application only against a background of 'normal circumstances' corresponding to the type of context in which the words were used in the past. There is no 'convention' to guide us as to whether or not a particular expression applies in some extraordinary situation. This is not because the meaning of the word is 'vague', but because the application of words ultimately depends on there being a sufficient similarity between the new situation of use and past situations. The relevant dimensions of similarity are not fixed once and for all; this is what generates 'open texture' (Waismann 1951). Ultimately, it is the context of utterance which determines which dimension of similarity is relevant, hence which conditions have to be satisfied for a given expression to apply (Travis 1975; 1981). It follows that the sense of ordinary descriptive words is context-dependent, like the reference of indexicals, though not quite in the same way. On this approach, which we may call 'contextualism', truth-conditions cannot be ascribed to sentence-types but only to utterances (Searle 1978; 1983).

Contextualism was a central tenet in the pragmatic conception of language developed by ordinary language philosophers (though some atypical ordinary language philosophers, like Grice, rejected it). This conception is, at bottom, a 'use theory of meaning'. Meaning is use, in the following sense: there is nothing more to meaning than use (Wittgenstein 1953). We are confronted with uses of words, and the meaning which those words acquire for us is only the sense we are able to *make* of those uses.

8 The semantics/pragmatics distinction

If much of contemporary pragmatics derives from the work of ordinary language philosophers, the name 'pragmatics' - contrasted with 'syntax' and 'semantics' - was coined by a philosopher in the ideal language tradition, Charles Morris (1938). The general conception associated with Morris' tripartite distinction has been very influential. On this conception (hereafter 'the traditional conception'), semantics and pragmatics are

complementary studies: semantics deals with meaning understood as representational content, whereas pragmatics deals with use. This view, still influential today, is at odds with the more radical conception developed within ordinary language philosophy and sketched in the previous sections. (According to the latter view, meaning cannot be divorced from use; semantics *is* pragmatics.) However, as we shall see, the border between semantics and pragmatics has become much fuzzier as the traditional conception, with its sharp contrast between the two disciplines, was modified in order to account for indexicality, pragmatic indications and related phenomena.

The traditional conception is commonly glossed in two different ways, corresponding to two different distinctions. The first distinction is that between 'meaning' (that is, representational content) and 'force', to be found both in Frege and Austin. (Often the term 'sense' is used instead of 'meaning', and the distinction referred to as the 'sense-force distinction'. Some commentators insist on distinguishing the use of 'sense' as the complement of 'force' from the use of 'sense' as the complement of 'reference'; see Sense and reference.) An utterance (1) represents a certain state of affairs, and (2) serves to perform a certain speech act. The semantics/pragmatics distinction is often expressed in terms of this distinction: semantics deals with representational content and studies the relations between words and the world, while pragmatics studies the relations between words and their users. The second distinction is that between sentence meaning and utterance meaning. Semantics is supposed to deal with the linguistic meaning of sentence-types, while pragmatics is concerned with the total significance of an utterance of the sentence by a particular speaker in a particular context. The traditional conception is indifferently expressed in terms of either distinction because, according to the traditional conception, the conventional meaning of the sentence-type *is* its representational content.

The problem with the traditional conception is precisely that it rests on the equation of the meaning of a sentence with its representational content. This equation cannot be accepted, for two reasons:

(1) Indexical sentences represent a specific state of affairs only in context; their representational content depends on some feature of their context of use, hence it cannot be equated with the context-independent meaning of the sentence-type. Because the truth-conditions of indexical sentences depend on their use, some authors have argued that the study of truth-conditions for such sentences belongs to pragmatics (Bar-Hillel 1954); it has even been suggested that pragmatics *is* the study of truth-conditions for indexical sentences (Montague 1968).
(2) As we have seen, there are use-conditional as well as truth-conditional aspects of the meaning of sentence-types. To account for them, it seems we must give up the purely truth-conditional conception of semantics and make room for a non-truth-conditional component within it. Alternatively, we can say that it is the business of 'pragmatics' to deal with some aspects of the linguistic meaning of sentence-types, namely those aspects which relate to use. Thus Gazdar (1979) defines pragmatics as 'semantics minus truth-conditions'.

Some philosophers have tried to defend the traditional conception by forcing use-conditional aspects of meaning into the mould of truth-conditional semantics. Take, for example, the imperative mood. One can use the pragmatic equivalence between the imperative 'Close the door!' and the 'explicit performative' 'I order you to close the door' (Austin 1975: 32) to support the claim that non-declarative sentences have a declarative paraphrase through which they can be given a truth-conditional analysis (Lewis 1970: 54-61). A number of similar attempts have been made to reduce use-conditional to truth-conditional aspects of meaning (for example, Davidson 1979). Despite these attempts, it is commonly acknowledged that not all aspects of linguistic meaning are truth-conditional. There are two components in the meaning of a sentence: a truth-conditional or descriptive component and a non-truth-conditional, pragmatic component. The pragmatic component of sentence meaning constrains the context of utterance: it is a 'procedural' component. As for the descriptive component of sentence meaning, it can no longer be equated with the truth-conditional content of the utterance (because of objection (1)), but it can still be construed as what *determines* that content with respect to a context of use.

This leaves us with two interpretations of the semantics/pragmatics distinction, both current in the contemporary literature. On one interpretation, dominant among philosophers, pragmatics deals with use - including use-conditional aspects of meaning (pragmatic indications, presuppositions and the like) - while semantics deals both with the descriptive component of sentence meaning and the truth-conditional content it determines (with respect to a context). The other interpretation, widespread among linguists, has it that semantics deals with conventional sentence meaning under all its aspects (including non-truth-conditional aspects), while pragmatics deals with use and the aspects of meaning which are contextual and use-dependent, that is, those that are conveyed

by the utterance but cannot be ascribed to the sentence-type (Katz 1977: 13- 22). (This category of 'contextual meaning' is somewhat heterogeneous; it includes both the semantic values of context-sensitive expressions, which are constitutive of the proposition literally expressed by the utterance, and other aspects of meaning which are not 'literal', conversational implicatures, for example.)

9 Context and propositional attitudes

The descriptive component of sentence meaning can be equated with a function from contexts to propositions (Stalnaker 1970; Kaplan 1989: 481-563). The 'context' is often construed as a package of various situational factors relevant to determining the semantic values of the context-sensitive constituents of the sentence. (See §10 for an alternative construal.) Thus the place of utterance, the identity of the participants in the speech episode and the time of utterance are among the factors on which the proposition expressed by an indexical sentence depends. It would be a mistake, however, to hold that only such 'external' (that is, non-intentional) features of the situation of utterance have a role to play in the determination of what is said. In many cases, what the speaker 'has in mind' is the relevant factor. Thus 'John's book' can mean the book which John wrote, the book he bought, the book he is reading, and so forth. The sentence in which the expression occurs expresses a definite proposition only when a particular relation between John and a certain book has been contextually determined, but there is no 'rule' which enables the interpreter to determine the latter except that it must be the relation which the speaker 'has in mind'. (This is in contrast with the case of 'I': as Barwise and Perry (1983: 33) pointed out, the reference of 'I' is fixed by the rule that 'I' refers to the speaker, irrespective of the speaker's beliefs and intentions. Even if a speaker believes that they are Napoleon, their use of 'I' does not refer to Napoleon.) The same thing holds for demonstratives in general: contrary to the received opinion, the reference of a demonstrative is not the 'demonstrated' object, for there may be no accompanying demonstration; the reference, rather, is the object which the speaker has in mind and wishes to single out (Kaplan 1989: 565-614; Bach 1987).

It turns out that the context against which an utterance is interpreted includes factors like the intentions, expectations, beliefs and other propositional attitudes of the speaker and their audience. Especially important are the beliefs which are shared and 'mutually known' to be shared; they constitute a 'common ground' which can be exploited in discourse (see Stalnaker's 1974 paper 'Pragmatic Presuppositions' in Davis 1991; Clark 1992). This introduces the topic of 'presupposition' which is generally mentioned, along with speech acts, indexicals and implicatures, as one of the central issues in pragmatics.

10 Presupposition

There is a basic sense in which 'presupposing' is a pragmatic attitude towards a proposition: that of 'taking it for granted'. (One takes something for granted, for example, when one uses it as a hidden premise in an argument.) The 'context' is sometimes defined as a set of presuppositions in this sense, that is a set of propositions which are taken for granted at a given point in discourse (see Karttunen 1974, for example). Many authors think that in order to be part of the context, a proposition must be not only believed by the participants in the speech episode, but also believed to be believed, and so forth. Other authors find this 'mutual belief' requirement too strong (Smith 1982). Be that as it may, a more pressing question arises in connection with presuppositions. Beside the pragmatic notion of presupposition (where presupposing is something a *speech participant* does), is there also a purely semantic notion, where presupposing is something which a *sentence* does? For example, does the sentence 'John stopped teaching undergraduates' carry the presupposition that John used to teach undergraduates as part of its semantic, truth-conditional content?

The semantic notion of presupposition has been questioned on two grounds. First, it has been pointed out that presuppositions, like conversational implicatures, seem to be defeasible or cancellable. This might suggest that the basic, pragmatic sense is the only sense we can give to the notion of presupposition. *Sentences* do not have presuppositions; only the participants in a speech episode can presuppose something. This conclusion, however, seems too strong, for the conventional nature of presuppositions is manifest and well-documented. Arguably, what the defeasibility of presuppositions shows is not that presuppositions are non-conventional, but rather that they can be overridden if certain conditions are satisfied (Gazdar 1979).

More convincing is the claim that presuppositions, though part of the conventional meaning of the sentence, do not affect the truth-conditions of the utterance. Like pragmatic indications in general, the linguistic presuppositions

associated with certain expressions (such as the verb 'stop' in the example above) can be construed as conditions of use or constraints on the context (see Stalnaker's 1974 paper 'Pragmatic Presuppositions' in Davis 1991). The linguistic presupposition encoded by the verb 'stop' is a certain constraint on the context, namely the requirement that it contain a speaker with a certain pragmatic attitude (the attitude of 'presupposing') towards a certain proposition, or (if the context is directly construed as a set of propositions) the requirement that it contain a certain proposition, namely the proposition that John used to teach undergraduates. An utterance of 'John stopped teaching undergraduates' is 'appropriate' only in a context in which this constraint is satisfied. The constraint in question belongs to the non-truth-conditional component of sentence meaning: it does not affect the (truth-conditional) 'content' of the utterance (see Presupposition).

11 Interpretation and context-change

The rich, propositional notion of 'context' which features in discussions of presupposition is at the heart of contemporary pragmatics. As we have seen in connection with demonstratives and semantically indeterminate expressions, the context provides 'assumptions' concerning the speaker's intentions and expectations, which are used to determine the proposition expressed by the sentence. This process of determination is construed as fundamentally inferential and proposition-involving. (See §14 on the intentional-inferential approach.)

The propositional notion of context makes it possible to see the relation between context and content as two-way rather than one-way (Kamp 1985: 240). The proposition expressed, which depends on the context, itself changes the context. According to dynamic theories of discourse, the content of an assertion is normally fed into the context against which the next utterance will be interpreted (Karttunen 1974; Stalnaker 1978; Kamp 1985; Heim's 1988 paper 'On the Projection Problem for Presuppositions' in Davis 1991). The context of interpretation constantly changes - Stalnaker speaks of an 'everchanging context' - because it evolves as discourse proceeds. Thus it is possible for the context to shift in the middle of an utterance. This possibility accounts for a number of puzzling facts, including the defeasibility of presuppositions (see Discourse semantics).

If the proposition expressed by an utterance is normally fed into the context, the assumption that this proposition has been expressed *always* becomes part of the context as a result of the interpretation of the utterance. It is this assumption, together with the default assumption that the speaker respects the norms of conversation (plus various other assumptions included in the context), which make it possible to infer the conversational implicatures which enrich the overall meaning of the utterance. It follows that the contextual changes induced by an utterance by virtue of its expressing a certain proposition affect not only the interpretation of the utterances that follow, but equally the overall meaning of the very utterance responsible for the contextual change.

Another sort of context-change induced by an utterance has been described by David Lewis (1979). Sometimes the default assumption that the speaker respects the norms of conversation prevents the utterance from being interpreted with respect to the context at hand because, if it were so interpreted, it would violate the norms in question. This leads to a modification of the context in order to reach a more satisfactory interpretation. Thus if the utterance presupposes that *P*, and *P* is not part of the context at hand, it is introduced into the context in order to bring the utterance into conformity with the norms ('accommodation').

12 The strategic importance of conversational implicatures

If semantics and pragmatics both study the contextual determination of the proposition expressed (in so far as it depends both on the linguistic meaning of the sentence and the context), conversational implicatures fall within the sole domain of pragmatics, for they are not constrained by the linguistic meaning of the sentence in the way the proposition expressed is. Yet the theory of implicatures has important consequences for semantics. Thanks to Grice's theory, many intuitive aspects of meaning can be put into the 'pragmatic wastebasket' as implicatures, rather than treated as genuine data for semantics. Take, for example, the sentence '*P* or *Q*'. It can receive an inclusive or an exclusive interpretation. Instead of saying that 'or' is ambiguous in English, we may consider it as unambiguously inclusive, and account for the exclusive reading by saying that in some contexts the utterance conversationally implicates that *P* and *Q* are not both true. When there is such a conversational implicature, the overall meaning of the utterance is clearly exclusive, even though what is strictly and literally said corresponds to the logical formula '$P \lor Q$'. It is here that the complementary character of semantics and pragmatics is particularly manifest. Semantics is simplified because a lot of data can be explained away as 'implicatures' rather

than genuine aspects of the (literal) meaning of the utterance.

Grice's theory of implicatures has been extremely popular among semanticists precisely because it enables the theorist, when certain conditions are satisfied, to shift the burden of explanation from semantics to pragmatics. From this point of view, the most interesting notion is that of 'generalized' conversational implicatures (Grice 1989; Gazdar 1979; Levinson 1983). When a conversational implicature is generalized, that is generated by default, it tends to become intuitively indistinguishable from semantic content. Grice's theory has taught the semanticist not to take such 'semantic' intuitions at face value. Even if something seems to be part of the semantic content of an utterance, the possibility of accounting for it pragmatically must always be considered.

Grice's theory is important also because it has provided an influential argument against the contextualism professed by ordinary language philosophers. For example, Strawson had claimed that the truth-conditions of '*P* and *Q*' in English are contextually variable: the notion of temporal succession, or that of causal connection, or a number of other suggestions concerning the connection between the first and the second conjunct can enter into the interpretation of '*P* and *Q*', depending on the context (Strawson 1952: 81-2). 'They got married and had many children' means that they had children after getting married; 'Socrates drank the hemlock and died' means that he died as result of drinking the hemlock. Those aspects of the interpretation are very much context-sensitive; yet they affect the utterance's truth-conditions. The truth-conditions of '*P* and *Q*', therefore, are not fixed by a rigid rule, but depend on the context. Against this view, Grice has argued that the truth-conditions of '*P* and *Q*' *are* fixed and context independent. '*P* and *Q*' is true if and only if *P* and *Q* are both true. Thus 'They got married and had many children' would be true, even if they had the children before getting married. Certainly the utterance conveys the suggestion that the children came after the marriage, but this suggestion is nothing other than a conversational implicature, according to Grice (see Implicature §6). It does not affect the utterance's semantic content - its literal truth-conditions. Grice criticized his fellow ordinary language philosophers for confusing the truth-conditions of an utterance with its total significance. Though controversial (Travis 1985), this argument has been very popular, and it has played a major role in the subsequent downfall of ordinary language philosophy.

13 Communicative intentions

The pioneers of pragmatics (Malinowski and Austin, for example) used to insist on the social dimension of language as opposed to its cognitive or representational function. As pragmatics developed, however, it is the psychological dimension of language use that came to the forefront of discussions, in part as a result of Grice's work on meaning and communication.

Grice ([1957] 1989) defined a pragmatic notion of meaning: the notion of *someone* meaning something by a piece of behaviour (a gesture, an utterance and so forth). Grice's idea was that this pragmatic notion of meaning was basic and could be used to analyse the semantic notion, that is what it is for a linguistic *expression* to have meaning (see Grice, H.P. §3). Strawson soon pointed out that Grice's pragmatic notion of meaning could also be used to characterize the elusive notion of an illocutionary act (Strawson [1964] 1971). In §2, the view that illocutionary acts are essentially conventional acts (like the acts which owe their existence to the rules of a particular game) was mentioned. This conventionalist approach was dominant in speech act theory until Strawson established a bridge between Grice's theory of meaning and Austin's theory of illocutionary acts. Illocutionary acts, in the new framework, can be analysed in terms of the utterly non-mysterious notion of a 'perlocutionary act'.

A perlocutionary act consists in bringing about certain effects by an utterance. For example, by saying to you 'It is raining', I bring it about that you believe that it is raining. Now, according to the suggested analysis, to perform the illocutionary act of asserting that it is raining is (in part) to make manifest to the addressee one's intention to bring it about, by this utterance, that the addressee believes that it is raining. An illocutionary act therefore involves the manifestation of a corresponding perlocutionary intention. But there is a special twist which the suggested analysis inherits from Grice's original conception of meaning: the intention must be made manifest in a specially 'overt' manner. Not only must the speaker's intention to bring about a certain belief in the addressee be revealed by his utterance, but his intention to reveal it must also be revealed, and it must be revealed in the same overt manner. This characteristic (if puzzling) feature of overtness is often captured by considering the revealed intention itself as reflexive: a communicative intention, that is the type of intention whose manifestation constitutes the performance of an illocutionary act, is the intention to achieve a certain perlocutionary effect (for example, bringing about a certain belief in the addressee) via the addressee's recognition of this intention. Also

relevant to the characterization of overtness is the notion of 'mutual knowledge' (Lewis 1969; Schiffer 1972) which we have seen at work in the characterization of 'contexts' and 'presuppositions' (see §10) (see Communication and intention; Meaning and communication).

14 The intentional-inferential model

Even though the conventionalist approach to communication is still alive, the Grice-Strawson 'intentionalist' approach has gained wide currency in pragmatics. Typical in this respect are the neo-Gricean theories offered by Bach and Harnish (1979) and Sperber and Wilson (1986). They have put forward an inferential model of communication intended to supersede the 'code model' (or 'message model') that was inspired from Shannon and Weaver.

According to the code model, communication proceeds as follows: to communicate a certain content, the speaker encodes it into a sentence using the grammar of the language as a 'code' pairing contents and sentences (possibly with respect to a context of utterance); the interpreter, by virtue of their knowledge of the same grammar (and, perhaps, of the context), can decode the sentence and recover the intended content.

The alternative, inferential model of communication is very different. An utterance is seen as a meaningful action, one which provides interpreters with evidence concerning the agent's intentions. What distinguishes communicative acts from other meaningful actions is what can be inferred from the evidence: a communicative act is an act which provides evidence of a certain 'communicative intention' on the part of the speaker. In other words, the speaker's intention to communicate something is what explains their utterance, when considered as a piece of behaviour. From this point of view, the content of the communicative act - what is communicated - is the total content of the communicative intentions which can be inferred from it. Let us call this the utterance's communicative meaning, distinct from the literal or conventional meaning of the sentence (determined by the code, that is the grammar). Understanding is essentially an inferential process in this framework, and the conventional meaning of the sentence provides only part of the evidence used in determining the communicative meaning of the utterance.

15 Pragmatics and modularity

A characteristic feature of recent work in the Gricean tradition has been the explicit employment of concepts from (and the intention to contribute to) cognitive science. For example, Fodor's distinction between central thought processes and more specialized cognitive 'modules' has been found relevant to the characterization of the task of pragmatics (see Modularity). (See, for example, Kasher's papers in Davis 1991 and Tsohatzidis 1995.)

In the inferential framework, comprehension involves not only a specifically linguistic competence, namely knowledge of 'grammar', but also general intelligence (that is, world knowledge together with inferential abilities). Contextual assumptions of various sorts, including assumptions about the speaker's beliefs and expectations, play a crucial role not only in the inferential process which, according to Grice and his followers, underlies the generation of conversational implicatures, but also in the determination of the proposition literally expressed by the utterance (Carston 1988; Kempson 1988). Pragmatics, therefore, is concerned with the interaction between the language faculty and central thought processes in the task of linguistic comprehension (Sperber and Wilson 1986). It considers how the output of the linguistic module is centrally processed, that is processed against the interpreter's complete belief system.

The major difficulty here is the 'holistic' character of the belief system (Fodor 1983). How is the total cognitive background restricted so as to yield a 'context' of manageable size? Sperber and Wilson's 'relevance theory' addresses this crucial issue. Their view of pragmatics is grounded in a general theory of cognition as relevance-oriented. One of their central ideas is that the context in which an utterance is processed is not 'given' but 'constructed': it results from an active search driven by the overarching goal to maximize relevance.

Sperber and Wilson's claim that there is no special faculty or module corresponding to pragmatics (since pragmatics studies the interaction between general intelligence and linguistic modules) must be qualified. As Sperber himself stressed in various places (for example, Sperber 1994), there is one special capacity which communicators must possess and which, arguably, poor communicators (for example, autistic children) do not possess: the capacity to ascribe complex propositional attitudes, such as higher-order intentions. (As we have seen,

understanding crucially involves a process of intention recognition, and the communicative intentions which must be recognized are complex, higher-order intentions.)

16 Cognitive science and contextualism

The intentional-inferential framework has been very influential in cognitive science. The Grice-Strawson interpretation of speech act theory has given rise to computational models of discourse where intention recognition plays a key role (see Grosz, Pollack and Sidner 1989; Cohen, Morgan and Pollack 1989). The idea that meaning is 'inferred' has also been taken very seriously. Many cognitive scientists believe that in interpretation the meaning of an utterance is arrived at by 'guesswork', rather than by decoding (Green 1989).

Typical in this respect are complex nominals (noun-noun compounds like 'finger cup' or adjective-noun compounds like 'philosophical kitchen'), which have received a good deal of attention. The semantic value which such expressions assume in context (unless they are idiomatic) is not predictable on a purely linguistic basis; it results from an act of 'sense creation'. Here again, what matters is what the speaker has in mind; the interpreter can only guess, using a variety of contextual clues. Such expressions whose sense is irreducibly contextual have been dubbed 'contextual expressions' (Clark 1992).

Contextual expressions are 'semantically indeterminate'. The extent of indeterminacy in language threatens the standard picture adhered to by most semanticists. According to that picture, 'what is said' is conventionally determined by the meaning of the sentence and the context. Then, and then only, is what the speaker really means 'inferred', if there is reason to think that what the speaker means is distinct from what is said (for example, if the speaker's literal contribution seems conversationally inappropriate and some further assumption is required in order to restore its conformity to conversational norms). As against this, both what is said and what is implied are seen as resulting from guesswork, because of semantic indeterminacy.

The fallback position for semanticists consists in holding that at least the meaning of the sentence-type is 'decoded' and can be read off the meanings of the constituents (as given in the lexicon) and the way they are put together. Decoding stops, and guesswork starts, only when we go from the meaning of the sentence to what the speaker means (this including both what is said and what is conversationally implied). But the fallback position itself has come under attack in the cognitive science literature. The very notion of the linguistic meaning of an expression-type has been questioned.

What is the linguistic meaning of a word? Is it possible to draw a line between 'pure' lexical knowledge and world knowledge? Arguably it is not (Langacker 1987: 154-66). For Langacker, the meaning of a word is a point of access into an essentially encyclopedic network. There is no distinction between 'dictionary' and 'encyclopedia'. A similar idea underlies Rumelhart's claim that interpretation is 'top down' from the bottom up (Rumelhart 1979). According to Rumelhart, what linguistic expressions do is evoke certain 'schemata' in memory. Those memorized schemata are part of our knowledge of the world. Interpretation, whether literal or non-literal, consists in finding evoked schemata which fit the situation the speaker seems to be talking about.

Langacker and Rumelhart reject not only the dictionary/encyclopedia distinction but also the literal/non-literal distinction as traditionally conceived. They insist that the same cognitive processes are involved in literal and (for example) metaphorical interpretations (see Gibbs 1994; Recanati 1995). A psychologist, Douglas Hintzman, has gone even farther, questioning the most basic distinction: that between the meaning of an expression-type and the contextual meaning of the token. Hintzman has developed a multiple-trace memory model in which the cognitive experiences associated with past tokens of a word interact with the present experience involving a new token of the word to yield the contextual 'meaning' of the latter (Hintzman 1986). Hintzman's model does not appeal to the notion of the literal meaning of the word-type. Words, as expression-types, do not have 'meanings' over and above the collection of token-experiences with which they are associated. The only meaning that words have is that which emerges in context.

Rejecting the three distinctions above - literal/non-literal, dictionary/encyclopedia and type-meaning/token-meaning - amounts to rejecting the semantics/pragmatics distinction. The 'eliminativist' approach to linguistic meaning developed within cognitive science constitutes a return to contextualism, the radical conception of pragmatics associated with ordinary language philosophy (§7). Pragmatics absorbs semantics in a contextualist framework.

See also: Strawson, P.F.; Wittgenstein, L.

<div align="right">FRANÇOIS RECANATI</div>

References and further reading

Alston, W. (1964) *Philosophy of Language*, Englewood Cliffs, NJ: Prentice Hall.(A textbook in the ordinary language tradition.)

Atlas, J. (1992) *Philosophy without Ambiguity: A Logico-Linguistic Essay*, Oxford: Clarendon Press.(On semantic indeterminacy.)

Austin, J.L. (1971) *Philosophical Papers*, Oxford: Clarendon Press, 2nd edn.(A posthumous collection of Austin's papers.)

__ (1975) *How to Do Things with Words*, Oxford: Clarendon Press, 2nd edn.(Austin's lectures on speech act theory, delivered at Harvard University in 1955.)

Bach, K. (1987) *Thought and Reference*, Oxford: Clarendon Press.(Discusses a number of pragmatic issues.)

Bach, K. and Harnish, M. (1979) *Linguistic Communication and Speech Acts*, Cambridge, MA: MIT Press. (Speech act theory within an intentional-inferential framework.)

Bar-Hillel, Y. (1954) 'Indexical Expressions', in *Aspects of Language*, Jerusalem: Magnes Press, 1970, 69-88. (Classic paper by a disciple of Carnap who was influenced by ordinary language philosophy. The 1970 collection also contains Bar-Hillel's pioneering papers on indexicality, context-dependence and pragmatic implication.)

Baron-Cohen, S., Leslie, A. and Frith, U. (1985) 'Does the autistic child have a "theory of mind"?', *Cognition* 21: 37-46.(See §15.)

Barwise, J. and Perry, J. (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press/Bradford Books.(A semantic theory which takes context-sensitivity seriously.)

Blakemore, D. (1992) *Understanding Utterances: An Introduction to Pragmatics*, Oxford: Blackwell.(An introduction to Relevance Theory.)

Bühler, K. (1990) *Theory of Language*, trans. D. Goodwin, Amsterdam: Benjamins.(Originally published in 1934, still worth studying.)

Carston, R. (1988) 'Implicature, Explicature, and Truth-Theoretic Semantics', in R. Kempson (ed.) *Mental Representations: The Interface Between Language and Reality*, Cambridge: Cambridge University Press, 155-81.(On levels of meaning.)

Clark, H. (1992) *Arenas of Language Use*, Chicago, IL: University of Chicago Press/CSLI.(An important collection of papers by a pragmatically-oriented psychologist, most of them written in collaboration. Part 1 concerns the context as 'common ground'; part 3 deals with 'sense creation'.)

Cohen, P., Morgan, J.L. and Pollack, M. (eds) (1989) *Intentions in Communication*, Cambridge: MA: MIT Press.(Intention recognition and speech acts from a computational point of view.)

Cole, P. (ed.) (1978) *Syntax and Semantics 9: Pragmatics*, New York: Academic Press.(Includes papers by Kaplan, Stalnaker and Grice, among others.)

Cole, P. (ed.) (1981) *Radical Pragmatics*, New York: Academic Press.(Several papers on the interface between semantics and pragmatics.)

Cole, P. and Morgan, J.L. (eds) (1975) *Syntax and Semantics 3: Speech Acts*, New York: Academic Press.(An influential collection of papers, mostly on 'indirect speech acts'; includes Grice's third William James Lecture, on conversational implicatures.)

Davidson, D. (1979) 'Moods and Performances', in Margalit, A. (ed.) *Meaning and Use*, Dordrecht: Reidel, 1979, 9-20.(How to force pragmatic indications into the mould of truth-conditional semantics.)

Davis, S. (ed.) (1991) *Pragmatics: A Reader*, New York: Oxford University Press.(An inclusive collection, containing many influential papers.)

Ducrot, O. (1972) *Dire et ne pas dire: principes de sémantique linguistique (Saying and Not Saying: Principles of Linguistic Semantics)*, Paris: Hermann.(A speech-act theoretic approach to presupposition; contains insightful discussions of many pragmatic issues.)

Fodor, J. (1983) *The Modularity of Mind*, Cambridge, MA: MIT Press/Bradford Books.(Distinguishes between input systems and central systems in the human mind.)

Gardiner, A. (1932) *The Theory of Speech and Language*, Oxford: Clarendon Press.(A remarkable anticipation of speech act theory. Still worth reading.)

Gazdar, G. (1979) *Pragmatics: Implicature, Presupposition and Logical Form*, New York: Academic Press. (Rigorous work in the formal-semantic tradition.)

Gibbs, R. (1994) *The Poetics of Mind*, Cambridge: Cambridge University Press.(Argues that the same cognitive processes are at work in literal and non-literal communication.)

Green, G. (1989) *Pragmatics and Natural Language Understanding*, Hillsdale, NJ: Erlbaum.(An introduction to pragmatics for cognitive scientists.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge: Cambridge University Press.(Contains the William James Lectures and a collection of Grice's papers on the philosophy of language, including 'Meaning'.)

Grosz, B., Pollack, M. and Sidner, C. (1989) 'Discourse', in M. Posner (ed.) *Foundations of Cognitive Science*, Cambridge, MA: Bradford Books/MIT Press, 437-68.(A survey of computational pragmatics.)

Heim, I. (1988) 'On the Projection Problem for Presuppositions', in S. Davis (ed.) *Pragmatics: A Reader*, New York: Oxford University Press, 1991, 397-405.(Uses the notion of 'context-change potential'.)

Harnish, R.M. (1976) 'Logical Form and Implicature', in S. Davis (ed.) *Pragmatics: A Reader*, New York: Oxford University Press, 1991, 316-64.(How to shift the burden of explanation from semantics to pragmatics.)

Hintzman, D. (1986) '"Schema Abstraction" in a Multiple-Trace Memory Model', *Psychological Review* 93 (4): 411-28.(An exemplar model of categorization which has deep implications for semantics and pragmatics.)

Horn, L. (1989) *A Natural History of Negation*, Chicago, IL: University of Chicago Press.(The last three chapters of this monograph deal with the interplay between semantic and pragmatic factors in interpretation.)

Kamp, H. (1985) 'Context, Thought and Communication', *Proceedings of the Aristotelian Society* 85: 239-61.(An informal presentation of Discourse Representation Theory.)

Kaplan, D. (1989) 'Demonstratives' and 'Afterthoughts', in J. Almog, H. Wettstein and J. Perry (eds) *Themes from Kaplan*, New York: Oxford University Press, 481-563, 565-614.('Demonstratives' deals with the logic of indexicals, with 'Afterthoughts' providing further comments.)

Karttunen, L. (1974) 'Presupposition and Linguistic Context', *Theoretical Linguistics* 1: 181-94.(The third in an important series of papers by Karttunen on presuppositions.)

Kasher, A. (ed) (1997) *Pragmatics: Critical Assessment*, London: Routledge.(A new reader, very inclusive and containing postscripts by some of the authors.)

Katz, J. (1977) *Propositional Structure and Illocutionary Force*, New York: Crowell.(Semantics, pragmatics and speech act theory.)

Kempson, R. (1988) 'Grammar and Conversational Principles', in F. Newmeyer (ed.) *Linguistics: The Cambridge Survey*, vol. 2, 139-63.(Redraws the boundary between semantics and pragmatics.)

Langacker, R. (1987) *Foundations of Cognitive Grammar*, vol. 1, Stanford, CA: Stanford University Press.(A new paradigm, based on the rejection of traditional distinctions, such as syntax/semantics and semantics/pragmatics.)

Levinson, S. (1983) *Pragmatics*, Cambridge: Cambridge University Press.(The standard textbook.)

Levinson, S. (1988) *Generalised Conversational Implicatures and the Semantics/Pragmatics Interface*, unpublished typescript.(How pragmatics affect truth-conditions.)

Lewis, D. (1969) *Convention*, Cambridge, MA: Harvard University Press.(Introduces the notion of 'common knowledge'.)

Lewis, D. (1970) 'General Semantics', *Synthese* 22: 18-67.(Section 8 provides a truth-conditional treatment of non-declaratives.)

Lewis, D. (1973) *Counterfactuals*, Oxford: Blackwell.(Introduces 'salience' in connection with incomplete definite descriptions.)

Lewis, D. (1979) 'Scorekeeping in a Language-Game', *Journal of Philosophical Logic* 8: 339-59.(On 'accommodation' and context-change.)

Lewis, D. (1980) 'Index, Context, and Content', in S. Kanger and S. Öhman (eds) *Philosophy and Grammar*, Dordrecht: Reidel, 79-100.(Shows the need for a distinction between 'index' and 'context'.)

Lyons, J. (1977) *Semantics*, Cambridge: Cambridge University Press.(Volume 2 of this textbook contains three chapters on pragmatic issues.)

Malinowski, B. (1923) 'The Problem of Meaning in Primitive Languages', supplement to C. Ogden and I.A. Richards, *The Meaning of Meaning*, London: Routledge, 10th edn, 1949, 296-336.(Speech as a mode of social action.)

Margalit, A. (ed.) (1979) *Meaning and Use*, Dordrecht: Reidel.(A collection of papers dedicated to the memory of Bar-Hillel, a pioneer of pragmatics.)

Montague, R.M. (1968) 'Pragmatics', in R. Klibansky (ed.) *La Philosophie Contemporaine 1*, Florence: La Nuova Italia Editrice, 102-22.(A classic paper in formal semantics.)

Morris, C. (1938) *Foundations of the Theory of Signs*, Chicago, IL: University of Chicago Press.(Introduces the famous trichotomy between syntax, semantics and pragmatics. See §8.)

Nerlich, B. and Clarke, D. (1994) 'Language, action and context: Linguistic pragmatics in Europe and America (1800-1950)', *Journal of Pragmatics* 22: 439-63.(The pre-history of pragmatics.)

Nunberg, G. (1979) 'The Non-Uniqueness of Semantic Solutions: Polysemy', *Linguistics and Philosophy* 3: 143-84.(Influential paper on the pragmatics of reference.)

Recanati, F. (1987) *Meaning and Force: The Pragmatics of Performative Utterances*, Cambridge: Cambridge University Press.(A survey of pragmatics and speech act theory, with special attention to the debate between conventionalist and intentionalist approaches.)

Recanati, F. (1993) *Direct Reference: From Language to Thought*, Oxford: Blackwell.(The second part deals with the demarcation between semantics and pragmatics.)

Recanati, F. (1995, forthcoming) 'The alleged priority of literal interpretation', *Cognitive Science*, forthcoming. (Criticizes Grice's sequential model of non-literal interpretation.)

Reddy, M. (1979) 'The Conduit Metaphor', in A. Ortony (ed.) *Metaphor and Thought*, Cambridge: Cambridge University Press, 2nd edn, 1994, 164-201.(A classic paper which has influenced Lakoff's theory of metaphor.)

Rumelhart, D.E. (1979) 'Some problems with the notion of literal meaning', in A. Ortony (ed.) *Metaphor and Thought*, Cambridge: Cambridge University Press, 2nd edn, 1994, 71-82.(An iconoclastic paper. See §16.)

Schiffer, S. (1972) *Meaning*, Oxford: Clarendon Press.(An important work in the Gricean tradition, introducing the notion of 'mutual knowledge'.)

Searle, J. (1969) *Speech Acts*, Cambridge: Cambridge University Press.(A systematic theory of speech acts within a conventionalist framework.)

Searle, J. (1978) 'Literal Meaning', in *Expression and Meaning*, Cambridge: Cambridge University Press, 1979, 117-36.(A collection of influential papers on speech act theory. 'Literal Meaning' generalizes context-sensitivity: words have conditions of satisfaction only against a background of assumptions and practices.)

Searle, J. (1983) *Intentionality*, Cambridge: Cambridge University Press.(Chapters 5 and 6 bear on issues such as literal meaning, the 'background', and the role of speaker's intentions.)

Simon, H. (1994) 'Literary Criticism: A Cognitive Approach', in G. Güzeldere and S. Franchi (eds) '"Bridging the Gap": Where Cognitive Science Meets Literary Criticism', *Stanford Humanities Review*, supplementary vol. 4 (1): 1-26.(A pragmatic theory of meaning by one of the founding fathers of cognitive science.)

Smith, N. (ed.) (1982) *Mutual Knowledge*, London: Academic Press.(Contains an interesting debate between Clark and Carlson and Sperber and Wilson over mutual knowledge.)

Sperber, D. (1994) 'Understanding Verbal Understanding', in J. Khalfa (ed.) *What is Intelligence?*, Cambridge: Cambridge University Press, 179-98.(The role of meta-representations in communication.)

Sperber, D. and Wilson, D. (1986) *Relevance: Communication and Cognition*, Oxford: Blackwell.(A cognitively oriented theory in the Gricean tradition. See §§14-15.)

Stalnaker, R. (1970) 'Pragmatics', *Synthese* 22: 272-89.(A lucid and original survey.)

Stalnaker, R. (1974) 'Pragmatic Presuppositions', repr. in S. Davis (ed.) *Pragmatics: A Reader*, New York: Oxford University Press, 1991, 471-82.(See §10.)

Stalnaker, R. (1978) 'Assertion', in P. Cole (ed.) *Syntax and Semantics 9: Pragmatics*, New York: Academic Press, 1978, 315-22.(Sketches a theory of speech acts in terms of context-change.)

Stenius, E. (1967) 'Mood and Language-Game', *Synthese* 17: 254-74.(Language as a rule-governed activity. See §4.)

Strawson, P. (1950) 'On Referring', in *Logico-Linguistic Papers*, London: Methuen, 1971, 1-27.(On presupposition and the sentence/statement distinction. The collections contains Strawson's most important papers in the philosophy of language.)

Strawson, P. (1952) *Introduction to Logical Theory*, London: Methuen.(Insists on the differences between natural and artificial languages.)

Strawson, P. (1964) 'Intention and Convention in Speech Acts', in *Logico-Linguistic Papers*, London: Methuen, 1971, 149-69.(A classic of speech act theory. See §13.)

Travis, C. (1975) *Saying and Understanding*, Oxford: Blackwell.(Emphasizes the context-dependence of sense.)

Travis, C. (1981) *The True and the False: the Domain of the Pragmatic*, Amsterdam: Benjamins.(A case for contextualism.)

Travis, C. (1985) 'On What is Strictly Speaking True', *Canadian Journal of Philosophy* 15: 187-229.(Criticizes Grice's critique of contextualism.)

Tsohatzidis, S.L. (ed.) (1995) *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, London: Routledge.(A new collection containing original papers on pragmatics and speech act theory.)

Waismann, F. (1951) 'Verifiability', in A. Flew (ed.) *Logic and Language*, 1st series, Oxford: Blackwell, 17-44.

(On 'open texture', by a disciple of Wittgenstein.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(A pragmatic approach to language.)

Predication

Some sentences have a very simple structure, consisting only of a part which serves to pick out a particular object and a part which says something about the object picked out. Expressions which can be used to say something about objects picked out are called predicates. Thus 'smokes' in 'Sam smokes' is a predicate. But 'predication' may refer either to the activity of predicating or to what is predicated. To understand either we need to know what predicates are and how they combine with other expressions.

Predicates, unlike proper names, can be negated. They combine with other expressions in ways described by categorial grammar: predicates are incomplete and are completed by other expressions, such as proper names. The word 'smokes' in 'Sam smokes' is called a monadic or 1-place predicate; predicates with two or more places are called relational predicates ('Sam loves Erna', '3 is between 2 and 4').

Since Frege it has been customary to hold that the incomplete or predicative parts of 'Sam smokes' and 'Sam is a smoker' are, respectively, 'smokes' and 'is a smoker' (see Frege, G. §§2-4). According to an older view, the incomplete part of 'Sam is a smoker' is the copula 'is', which is completed by two complete expressions, 'Sam' and 'a smoker' ('a' can be ignored as an accident of English). Sometimes called 'the two name theory of predication', this view allows two types of name - proper names and common nouns.

Is predication merely a matter of words? Are mental acts or the senses of words essential to it? For Frege, names and predicates are correlated with ideal senses, and the sentential wholes to which they belong with ideal thoughts or propositions (see Sense and reference §2; Frege 1892a). The sense or meaning of a proper name is sometimes called an individual concept, that of a predicate a concept (Frege himself used neither 'concept' nor 'meaning' in this way). It has been held that to use a predicate is to perform a mental act of predicating and that to use a proper name is to perform a mental act of naming or referring. Husserl and others appeal to both senses and mental acts and conceive of the latter as tokenings of the former. The unity of thoughts and of thinkings, on these views, resembles but is prior to the unity of the sentence: predicative senses require non-predicative senses; predicatings require referrings or acts of quantifying over individuals.

Predication occurs in the context of a variety of mental acts and linguistic actions: I may wonder or ask whether Sam smokes. So to predicate is not always to judge or assert. To assert that if Sam smokes he will smell is not thereby to assert that Sam smokes, although it is to predicate this of him in the context of entertaining or supposing the thought that Sam smokes.

Frege and Husserl pointed out that to deny that Sam smokes is just to assert that he does not smoke. Denying and judging are not on a par because 'not' belongs within the sentence, its sense to that of the sentence. There is, it is true, the phenomenon of polemic negation, as when the third word of the following sentence is stressed: 'Wittgenstein was not German'. But this is a pragmatic phenomenon.

What is it that is predicated? Just as names designate things, so too predicates are said to stand in a semantic relation (reference, signification) to one or more of the following: properties (attributes); what Frege (in his own technical sense) called 'unsaturated concepts'; sets; and relations between objects and possible worlds (see Semantics, possible worlds §9). Thus predication may be held to involve saying of Sam that he falls under the sense of the predicate 'smokes', exemplifies the property of being a smoker, or belongs to the set of smokers - or some combination of these. The semantic values of predicates and other parts of a sentence are sometimes said to form states of affairs or situations - which may or may not obtain - or facts.

The foregoing contains five simplifications and questionable presuppositions. First, does predication only occur within sentences (thoughts, thinkings)? If a builder shouts 'Slab!' to his assistant is he not producing a non-elliptic predication? Second, we may wonder, with Wittgenstein (1968), to whom the preceding example is due, whether there is in fact any uniform type of semantic relation between predicates, on the one hand, and properties and their ilk, on the other hand.

Third, the apparent (surface, grammatical) forms of predicates can diverge in different ways from their logical (deep) forms. To say that Pierre is tall is not to predicate the monadic property of being tall of him but to predicate,

of him and the average member of some reference class, the relational property of being taller than. Furthermore, if we accept the view (of Davidson; and of Parsons 1990) that many sentences involve quantification over events, then to say that Sam is smoking is to predicate of some event that it is an event of smoking and to predicate of it and Sam that Sam is the agent of it (see Adverbs §1).

Fourth, according to Ramsey (1990) there is no essential distinction, in 'Socrates is wise', between its subject and monadic predicate because the sentence expresses the same proposition as 'Wisdom is a characteristic of Socrates'. So no fundamental classification between individuals and properties can be based upon such a distinction.

Finally, if we take seriously the claim that predicates are incomplete, then we should say that in 'Sam smokes' the predicate is not 'smokes' but ' smokes', the gappy result of deleting the proper name from the sentence. Thus both 'Booth shot Lincoln' and 'Booth shot Booth' contain the predicate 'Booth shot '. But, as Geach indicates in his discussion of this example ('Quine on Classes and Properties', 1981: 222-5), the predicate here is not a physical part of the two sentences; it is a sentence pattern which 'Booth shot' helps to form. A related view is that such sentence patterns are abstract features of sentences, as melodies are of spoken sentences. But then we might wonder whether 'Booth ' is not just as much a sentence pattern as 'Booth shot '.

See also: Logical and mathematical terms, glossary of; Proper Names; Property theory; Strawson, P.F. §5

KEVIN MULLIGAN

References and further reading

Frege, G. (1891) *Funktion und Begriff*, Jena: Pohle; trans. P.T. Geach, 'Function and Concept', in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984, 137-56.(Classic account of the role of functions in predication.)

Frege, G. (1892a) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Meaning', in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984, 157-77.(The classic account of the sense and reference of sentences and their parts.)

Frege, G. (1892b) 'Über Begriff und Gegenstand', *Vierteljahrsschrift für wissenschaftliche Philosophie* 16: 192-205; trans. P.T. Geach, 'On Concept and Object', in *Collected Papers on Mathematics, Logic and Philosophy*, ed. B. McGuinness, Oxford: Blackwell, 1984, 182-94.(Classic account of what predicates and names signify.)

Geach, P.T. (1975) 'Names and Identity', in S. Guttenplan (ed.) *Mind and Language*, Oxford: Clarendon Press, 139-58.(Important treatment of the nature of predication, drawing on Frege.)

Geach, P.T. (1981) *Logic Matters*, Oxford: Blackwell.(See 'History of the Corruptions of Logic' for a historical sketch of the relations between the two name theory and the theory fully worked out by Frege.)

Gardies, J.-L. (1975) *Esquisse d'une grammaire pure*, Paris: Vrin; trans. K. Mulligan, *Rational Grammar*, Munich: Philosophia, 1985.(Useful guide to categorial grammar and its applications to natural languages.)

Parsons, T. (1990) *Events in the Semantics of English*, Cambridge MA: MIT Press.(Elegantly defends the controversial view that many sentences involve quantification over and predication of events and states.)

Ramsey, F. (1990) 'Universals', in *Philosophical Papers*, ed. H. Mellor, Cambridge: Cambridge University Press, 8-30.(Stimulating, critical examination of some ways of distinguishing between subjects and predicates and between particulars and properties.)

Simons, P. (1981) 'Unsaturatedness', *Grazer Philosophische Studien*, 14: 73-95.(Surveys different types of unsaturatedness and incompleteness and their relations to the notions of function, functor and predication.)

Strawson, P.F. (1974) *Subject and Predicate in Logic and Grammar*, London: Methuen.(Surveys fully most aspects of predication.)

Wittgenstein, L. (1968) *Philosophical Investigations*, Oxford: Blackwell.(Includes discussion of a series of examples that throw doubt on the plausibility of many common assumptions about sentences and semantic relations.)

Presupposition

There are various senses in which one statement may be said to 'presuppose' another, senses which are in permanent danger of being confused. Prominent among them are Strawsonian presupposition, a relation which obtains between statements when the falsity of one deprives the other of truth-value (for example, 'There was such a person as Kepler' is a Strawsonian presupposition of 'Kepler died in misery'); semantic presupposition, which obtains between a statement and a particular use of a sentence type, when the falsity of the statement means that that use will not after all constitute the making of a statement (for example, 'The name "Kepler" has a bearer' is a semantic presupposition of 'Kepler died in misery'); and pragmatic presupposition, a broader notion exemplified by the legitimate presumption that accepting or denying the statement 'Fred knows that the earth moves' means accepting 'The earth moves'.

1 Introduction

In most conversations a certain amount of knowledge is assumed by the parties. Conversing with people whom we know well, we can take it for granted that they know who we are and what we have recently been doing, along with much else. Ordinarily, 'presuppositions' are nothing other than these assumed parts of the conversational background.

The notions of presupposition which have been of greatest concern to linguists and philosophers, however, have been more narrowly defined (but see §4 below), typically being concerned with the way certain phrases or constructions indicate that the speaker is making a presupposition in the ordinary sense. So, if I affirm 'Wellington's victory at Waterloo was his greatest triumph', my form of words indicates that I am presupposing that Wellington was the victor at Waterloo and asserting that it was his greatest triumph. However, the commonly stressed opposition between presupposition and assertion is in some measure misleading. In asking 'Is Tony going to Salzburg again this summer?', I presuppose that he has been there before and inquire whether he intends to go there in the future; no assertion has been made.

This phenomenon may be closely related to another. If I affirm

(1) All John's children are asleep,

I presuppose that John has children and assert that they are asleep. In affirming

(2) If all John's children are asleep, we should not leave the house,

I do not make the same assertion, but I do still presuppose that John has children. It is this that linguists have in mind when they speak of presuppositions being easily 'inherited': they are often (but not always) retained even when the clauses that convey them are sententially embedded so as to lose their own assertive force. Much of the work done by linguists on this topic has been concerned with identifying constructions in which presuppositions do not survive, and to explain why inheritance is blocked (see various authors collected in Davies 1991).

Philosophers writing on the subject, by contrast, have largely been concerned to elucidate the nature of presupposition: to explain what one thing's presupposing another really amounts to; and to specify the kinds of item between which the relation of presupposition obtains.

2 Strawsonian presupposition

The writer most responsible for bringing the notion of presupposition to the attention of philosophers is P.F. Strawson, from whose work (1) above is drawn. Commenting on the situation of somebody who makes an assertion by uttering (1), Strawson remarks that

> he will not normally, or properly, say this, unless he believes that John has children…. But suppose…John has no children. Then is it true or false that all John's children are asleep? Either answer would seem to be misleading. But we are not compelled to give either answer. We can, and normally should, say that, since John has no children, the question does not arise.
>
> (1952: 173-4)

Strawson is eventually led to advance the following general explanation: 'if [statement] *S′* is a necessary condition of the *truth or falsity* of [statement] *S*,…then *S presupposes S*Čprime;' (1952: 175; original emphasis). So the statement that John has children is a 'Strawsonian presupposition' of the statement that all John's children are asleep; for unless the former is true the latter is neither true nor false. Strawson's talk of a 'question [which] does not arise' is liable to mislead. If John has no children, then the question 'Is statement (1) true or false?' may be raised; but the correct answer will be that it is neither.

To understand the notion of Strawsonian presupposition properly, it is necessary to understand what exactly a statement is. Strawson distinguishes between declarative 'sentences' - linguistic types the use of which on various occasions of utterance enables people to say things - and the 'statements' thereby made. It is statements, not sentences, that are true or false. Should a presupposition fail, there will be a 'truth-value gap': a statement will qualify neither as true nor false. An analogous notion applies to other kinds of speech act. So a yes/no question *Q* may be said to presuppose a statement *S* if the truth of *S* is a necessary condition of *Q*'s being answered by 'yes' or 'no' (see Strawson 1954).

The notion of Strawsonian presupposition has certainly found important philosophical application. In Strawson's own writings it forms the basis of a counter to Russell's suggestion that simple sentences involving 'definite descriptions' - that is, sentences of the form 'The *F* is *G*' - are not of subject-predicate form (see Descriptions). Russell held that such a sentence is best analysed as an existential sentence, 'There is one and only one *F*, and it is *G*', whereby the statement that the *F* is *G* straightforwardly entails the existence of an *F*. Having attacked Russell's arguments in its favour, it was on precisely this score that Strawson tried to refute the theory itself: he insisted that the statement that the present king of France is bald presupposes (rather than entails) that there is a present king of France (see Descriptions §4).

In assessing this suggestion, note that when contemplating statements which are neither true nor false, Strawson requires 'false' to bear a sense other than 'not true'. This means that, in the first place, the difference between Strawson and Russell is less than might at first have been supposed. Since Strawson accepts Russell's account of the conditions under which a statement made by uttering 'The *F* is *G*' will be true, he agrees *ipso facto* with his account of the conditions under which such a statement is not true. They differ only over the conditions for such a statement to be false, Strawson deeming such statements to be neither true nor false in circumstances where Russell would deem them straightforwardly false.

In the second place, however, we are driven to wonder what Strawson means by calling a statement 'false', given that it is not a synonym for 'not true'. It is reasonable to expect the attribution of falsity to a statement to be regulated by the principle that *S* is false just in case the negation of *S* is true, and this is a principle that Strawson's treatment respects; indeed, it seems that he takes this principle to be explanatory of the notion of falsity. But then it becomes crucial to have a test for identifying the negation of a statement, especially for natural languages, where the syntax does not necessarily make this obvious. Since Strawson's reluctance to classify 'The king of France is bald' as false when there is no king of France stems from a desire to reserve that attribution for the case in which there is a king of France and he is not bald, it is plain that Strawson regards the latter case as the proper negation of this statement. This, indeed, is what one might expect, given that Strawson is defending the view that 'The *F* is *G*' is a subject-predicate sentence. But the Russellian will want to draw a semantic distinction between 'It is not the case that the king of France is bald' and 'The king of France is not bald'. This suggests that the issue between them ultimately turns, not upon questions concerning presupposition, but upon the proper treatment of negation (and other logical modifiers) in natural languages.

3 Semantic presupposition

A rather different notion of presupposition is suggested by Frege: 'If anything is asserted, there is always an obvious presupposition that the simple or compound proper names used have a reference [*Bedeutung*]. If therefore one asserts "Kepler died in misery", there is a presupposition that the name "Kepler" designates something' (1892: 40). In order to see the difference between this notion of presupposition and Strawson's, it helps to make one assumption about statements; namely, that the same statement can be made in a variety of languages. Suppose, then, that there is a sequence of Chinese characters by the inscription of which it may be stated that Kepler died in misery. Such an inscription will certainly not contain the name 'Kepler', so it will not impede the making of that

statement, should the English (or German) name 'Kepler' lack a reference. But in that case,

(3) The name 'Kepler' has a reference

plainly is not a Strawsonian presupposition of

(4) Kepler died in misery,

that is, of the statement made by uttering (4). For the latter statement could be made by inscribing the Chinese characters instead, and the statement thereby made could have a truth-value even if (3) were false. What does qualify as a Strawsonian presupposition of (4) is the following statement:

(5) There was such a person as Kepler.

We need, then, to distinguish the new notion of presupposition, exemplified by (4) and (3), from that exemplified by (4) and (5).

In explicating this new notion, it helps to invoke again Strawson's distinction between a (declarative) sentence type - which is the bearer of meaning - and the statement made when the sentence is uttered or used - which is the bearer of truth-value. For with that distinction in mind, we see that the new notion is not a relation between statements, but a relation between a sentence - as used (with a particular meaning) upon a particular occasion - and a statement. Thus, the sentence 'You are drunk' (as used on a particular occasion) presupposes that the speaker is addressing somebody, in the sense that if not, then the use of that sentence on a particular occasion will not constitute the making of a statement. Generally, then:

A use of sentence σ in circumstance c presupposes statement S in so far as the truth of S is necessary for that use to qualify as the making of a statement.

When S and σ are so related, we may say that S is a 'semantic presupposition' of sentence σ, relative to c.

4 Pragmatic presupposition

Whatever the differences between them, Strawsonian and semantic presupposition are alike in depending upon the content of statements, or the meanings of sentence types. That is to say, they are alike in abstracting from the particular conversational purposes of the person uttering the sentence. In this respect they differ from another presupposition relation in which the actual speaker looms much larger. Following Stalnaker (1974), we might say that

A statement S' is a 'pragmatic presupposition' of a statement S in so far as a hearer may reasonably infer that the speaker accepts S' either from their acceptance of S or from their denial of it.

One salient difference between this explication and the definition of a Strawsonian presupposition is that the notions of truth and falsity have been replaced by those of acceptance and denial (acceptance of the negation). Another is the introduction of the notion of a hearer's reasonable inference - reasonable, that is, by the lights of a well-run conversation.

Whether one statement is a pragmatic presupposition of another depends crucially, then, upon what constitutes a well-run conversation, a notion philosophers have tried to explicate by formulating general 'conversational maxims', such as the rule (R) that one's hearer may assume that one is making the logically strongest relevant statement compatible with one's beliefs. To illustrate this, and to show how a pragmatic presupposition might obtain even when a Strawsonian presupposition does not, suppose that speakers are known to abide by (R), and consider the statements S: 'Fred knows that the earth moves' and S': 'The earth moves'. S' is not a Strawsonian presupposition of S: when S' is false, we may conclude simply that Fred's belief is not in fact knowledge and that S itself is straightforwardly false. All the same, a strong case can be made for the contention that S' is a pragmatic presupposition of S: I may reasonably infer that a speaker accepts S', should I hear them either accepting or denying S (see Stalnaker 1974). The case of accepting S presents no difficulty; and a denial of S amounts to an assertion of 'Fred doesn't know that the earth moves', an assertion which I may suppose that the speaker would not have made had they accepted the stronger statement 'The earth does not move'.

See also: Descriptions; Implicature; Logical and mathematical terms, glossary of; Pragmatics

IAN RUMFITT

References and further reading

Davies, S. (1991) *Pragmatics: A Reader*, New York: Oxford University Press. (Includes many papers showing the nature of contemporary work on presupposition by linguists, including articles on the problem of 'inheritance' mentioned in §1.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. 'On Sense and Meaning', in *Translations from the Philosophical Writings of Gottlob Frege*, trans. and ed. P.T. Geach and M. Black, Oxford: Blackwell, 3rd edn, 1980. (Introduces the notion of semantic presupposition. The page numbers of the original publication appear in the margins.)

Stalnaker, R.C. (1974) 'Pragmatic Presupposition', repr. in S. Davies, *Pragmatics: A Reader*, New York: Oxford University Press, 1991.

Strawson, P.F. (1950) 'On Referring', *Mind* 59: 320-44; repr. in *Logico-Linguistic Papers*, London: Methuen, 1971. (A classic paper, using presuppositional considerations to attempt to refute Russell's theory of descriptions. Note that the term 'presupposition' is not used in the paper; it is called 'implication in the special sense'.)

Strawson, P.F. (1952) *Introduction to Logical Theory*, London: Methuen. (Defines (Strawsonian) 'presupposition'.)

Strawson, P.F. (1954) 'A Reply to Mr. Sellars', *Philosophical Review* 63: 216-31. (Extends presupposition beyond the assertoric.)

Private language argument

Ludwig Wittgenstein argued against the possibility of a private language in his 1953 book Philosophical Investigations, where the notion is outlined at §243: 'The words of this language are to refer to what can be known only to the speaker; to his immediate, private, sensations. So another cannot understand the language.' The idea attacked is thus of a language in principle incomprehensible to more than one person because the things which define its vocabulary are necessarily inaccessible to others; cases such as personal codes where the lack of common understanding could be remedied are hence irrelevant.

Wittgenstein's attack, now known as the private language argument (although just one of many considerations he deploys on the topic), is important because the possibility of a private language is arguably an unformulated presupposition of standard theory of knowledge, metaphysics and philosophy of mind from Descartes to much of the cognitive science of the late twentieth century.

The essence of the argument is simple. It is that a language in principle unintelligible to anyone but its user would necessarily be unintelligible to the user also, because no meanings could be established for its signs. But, because of the difficulty of Wittgenstein's text and the tendency of philosophers to read into it their own concerns and assumptions, there has been extensive and fundamental disagreement over the details, significance and even intended conclusion of the argument. Some, thinking it obvious that sensations are private, have supposed that the argument is meant to show that we cannot talk about them; some that it commits Wittgenstein to behaviourism; some that the argument, self-defeatingly, condemns public discourse as well; some that its conclusion is that language is necessarily social in a strong sense, that is, not merely potentially but actually. Much of the secondary (especially the older) literature is devoted to disputes over these matters.

An account of the argument by the influential American philosopher Saul Kripke has spurred a semi-autonomous discussion of it. But Kripke's version involves significant departures from the original and relies on unargued assumptions of a kind Wittgenstein rejected in his own treatment of the topic.

1 The significance of the argument

The idea that a language might be private is explicitly canvassed in the second of Bertrand Russell's lectures 'The Philosophy of Logical Atomism', and Wittgenstein's argument, though possibly directed against his own earlier views in *Philosophical Remarks*, may originate in the sustained criticism of Russell which informs much of his writing.

Immediately prior to the *Philosophical Investigations*' discussion of private language, Wittgenstein suggests that the existence of the rules governing the use of words and making communication possible depends on agreement in human behaviour. This agreement includes people's reacting in similar ways to similar training, and exemplifies contingent 'very general facts of nature' which make particular concepts and customs possible and useful. Thus, for instance, one can train most children to look at something by pointing at it (whereas dogs look only at one's hand), and this enables us to attach meaning to the gesture of pointing and derivatives like signposts. The immediate function of the private language argument is to show that the possibility of linguistic rules and concept formation in general depends upon the possibility of such agreement.

But it has a further, connected, function. Motivating Wittgenstein's discussions of both mathematics and psychology is hostility to metaphysical absolutes, to the idea that we can find the world as it really is in the sense that any other way of conceiving it must be wrong (compare Wittgenstein 1953: 230). Both numbers and sensations provide especially tempting cases for philosophers, who are inclined to imagine that these are objects which force their identities upon us, our classifications of them and the rules governing the uses of their names being dictated to us in advance by the phenomena themselves. Wittgenstein makes precisely analogous points in his treatment of the two cases. In both, the underlying confusion is about how the act of meaning determines the future application of a formula or name.

With numbers, one temptation is to confuse the mathematical sense of 'determine' in which, say, the formula $y = 2x$ determines the numerical value of y for a given value of x (in contrast with $y (2x$, which does not) with a causal sense in which a certain training in mathematics determines that normal people will always write the same value

for *y* given both the first formula and a value for *x* (in contrast with creatures for whom such training might produce a variety of outcomes). This confusion produces the illusion that the outcome of an actual properly conducted calculation is the inevitable result of the mathematical determining, as though the formula's meaning itself were shaping the course of events.

In the case of sensations, the parallel temptation is to suppose that their natures are self-intimating - I seem to feel what they are directly, so that I need only give a name to one of them and the rules for the subsequent use of that name will be fixed on the spot. Wittgenstein tries to show that this is an illusion, that even that apparently most self-intimating of all sensations, pain, derives its identity only from a sharable practice of expression, reaction and use of language. Now if, say, pain were to force its identity upon me in the way described, then the possibility of such a shared practice, which in turn depends upon those general background facts of nature, would be irrelevant to the concept of pain. That is, if the real nature of part of the world were revealed to me in a single mental act of naming, as the private linguist supposes, then all subsequent facts would necessarily be irrelevant and the name could be private. The private language argument concludes that they could not be irrelevant, that no names could be private, and that the notion of having the real nature of the world so revealed is confused.

Such confusions underlie a range of articulated philosophical ideas and theories, without themselves being so articulated. The argument thus attempts, not to refute any particular theory, but to remove the motivation for involvement in a range of seemingly independent tasks, problems and solutions. Here are some frequently cited examples.

A still very common idea, found in Locke, is that interpersonal spoken communication works by speakers' translation of their internal mental vocabularies into sounds followed by hearers' re-translation into their own internal vocabularies (see Locke, J. §5). Again, Descartes considered himself able to talk to himself about his experiences while claiming to be justified in saying that he does not know (or not until he has produced a reassuring philosophical argument) anything at all about an external world conceived as something independent of them. And he and others have thought: while I may be wrong in my judgments about the external world, I am infallibly correct if they are restricted to my sensations. Again, many philosophers have supposed there to be a problem of other minds, according to which I may reasonably doubt the legitimacy of applying, say, sensation-words to beings other than myself (see Other minds). In all such cases, the implication is that my language could in principle be private: for these problems and theories even to make sense, the ability to share must be irrelevant to meaning and it must be conceivable that I am confined to my own case. (This suggestion is controversial: the usual charge is that such philosophers are committed to supposing that there actually is a private language. But this is not obvious, and the argument would be just as significant if all they are committed to is the supposition that their presupposed internal language, for all the difference it would make to its usability for self-communing, could be private.) This is especially clear in the case of Descartes, who must hold it possible to identify one's experiences inwardly, that is, without using any resources supplied by one's embodiment in an independently existing world, such as the concepts acquired in a normal upbringing. How is this identification to be achieved? This is the question considered in the next section.

2 The nature of the argument

The private language argument is usually identified with §§256-71 of *Philosophical Investigations*, a tightly-knit discussion of sensations, with §258 being especially important. But this discussion cannot properly be detached from the earlier sections of the book, despite commentators having frequently done just this. As important as the already-noted connection with the lengthy prior discussion of rule-following is the treatment of ostensive definition.

The argument proper is preceded by a preliminary discussion (§§244-55) in which Wittgenstein distinguishes two senses of 'private' and argues that natural languages (English, for example) are private in neither. The question then arises, could anything be a private language? He approaches this question by considering how a private language might be arrived at. It cannot be reached via the actual language we speak, for that is not private, and the attempt to convert it by thought experiment into a private one by simply suspending all expression of sensation and imagining the speaker just to name a sensation, as it were in a vacuum, merely raises the question of what this is supposed to consist in and what it is for. But to give private language enthusiasts a run for their money, Wittgenstein imagines himself in the position of establishing a private language for the purpose of keeping a

record of his sensations.

But he faces a serious impediment to discussing the matter at all: that doing so requires the mention of actions like ostensive definition, concentrating the attention, speaking, writing, remembering, believing, and so on, in the very process of suggesting that none of these can really be done in the circumstances under consideration. (This must be remembered in reading what follows, which in strictness should be constantly disfigured with scare quotes.)

He considers the idea that I simply *associate* a sign, say 'S', with a sensation by concentrating my attention on the sensation and saying 'S' to myself (the private analogue of ostensive definition), and points out that if this is to be a genuine definition it must establish a persisting connection between sign 'S' and that sensation: '"I impress [the connection] on myself" can only mean: this process brings it about that I remember the connection right in the future (§258)'.

This single remark has caused much trouble. Many have thought 'I remember the connection *right*' means 'I use 'S' only when I have S.' This had the argument resting on scepticism concerning memory, and provoked the criticism that memory's fallibility is neither more nor less a problem for a private linguist than a public one, so that the argument threatens both or neither. Wittgenstein's defenders retorted that fallibility is no problem only where mistakes can be corrected; in the private case, there can be no checking of memories, hence no chance of correction, and consequently talk of correctness is inappropriate. The critics responded by arguing either that such checking is possible, or that correctness does not require checkability.

This interplay of criticism and defence characterizes much of the commentary on the argument. But it is beside the point. Both sides mistakenly assume the connection in question between 'S' and the sensation to be a connection of *truth*, so that remembering the connection right is a matter of making the judgment that I am experiencing S only in the presence of S. But the connection Wittgenstein says must be remembered is a connection of *meaning*. It can now be seen that the argument makes no appeal to the fallibility of memory.

Imagine I am a private linguist. I have a sensation, and make the mark 'S' at the same time, as one might in an ordinary case introduce a sign by ostensive definition. Now suppose that later, I use 'S' in judging that I am again experiencing the same sensation. What do I mean by 'S' on this second occasion?

It cannot be that I mean the sensation I am now experiencing, for this collapses the distinction between meaning and truth and removes the status of factual assertion from my judgment that I am experiencing S - it becomes at best a fresh ostensive definition. (Thus I cannot say informatively that my shirt is orange and with the same utterance explain what 'orange' means by exhibiting my shirt.)

Can it be rather that I mean the sensation I named 'S' on the previous occasion? That would presuppose that I had indeed succeeded in giving my sensation the name 'S', but this cannot just be assumed: for there to have been an ostensive definition of 'S', a technique for the use of 'S' must have been established, one which leads to my using 'S' in the same way as before. This, though, is just what is in question. What would using the sign in the same way be here? The same way as what? Wittgenstein's earlier discussion of ostensive definition showed that there can be such definitions only where a place for them is already prepared - just pointing and making noises does not establish what sort of thing is being pointed at, even in the public world, and as a private linguist I cannot even do that but at best can only concentrate my attention. 'S' must be the name of a *sensation*, and the fact that I have had a sensation and simultaneously inwardly muttered 'S' does not suffice to make 'S' the name of that sensation. (To name a sensation is to name a *kind*, but what kind? A sensation may be classified in indefinitely many ways: ache, sensation-in-leg, sensation-at-time-*t*,….) And if one thinks a private linguist could remember the meaning of 'S' by remembering rightly the past correlation of 'S' with a certain kind of sensation, one presupposes what needs establishing: that there was such an independent correlation to be remembered. Fallibility of memory, even of memory of meaning, is neither here nor there: the point is that there has to be the right sort of occurrence in the first place to be a candidate for being remembered; and if there is not, no memory is going to create it. If, alternatively, we do not suppose that there is something to be remembered which is independent of the memory, then 'what seems right to me is right', that is, there is nothing to be right or wrong about. (These points are made in Wittgenstein 1953: 258-68.)

The conclusion is that it is impossible for a private linguist to establish and maintain a rule for the use of an expression, so that meaning is unobtainable in a private language. Only operating in a world independent of one's

impressions of it, in which one's operations are thus in principle available for scrutiny, can provide the possibility of real correlations of signs with objects and consistency in the usage of those signs.

3 Related considerations

The argument as given is embedded in a collection of arguments, observations and reminders on the connected topics of the relations between expression and description of experiences, between mind and behaviour, and between self-knowledge and knowledge of others, concentrated mainly in §§243-315 of *Philosophical Investigations* but spreading into other sections and other works. This whole collection is sometimes referred to as the private language argument. But the question of private language is present in several of Wittgenstein's even apparently unrelated discussions, such as those of seeing-as (seeing a triangle as resting on its base, for example), and kinaesthesis, where the only available descriptions of the experiences involved which do justice to their contents make direct reference to a sharable world of public objects.

Central to his thought here is the belief that no one argument will suffice to eliminate the commitment to private language, for it is the result of various powerful illusions. One of these is the idea that one can study the nature of mental phenomena by introspection, as though this gave us direct and unmediated access to the truth-condition for, say, 'My neck is itching', a condition to which only I am privy (see Introspection, psychology of). In contrast, Wittgenstein argues, the nature of mental phenomena is grasped not by introspection but by examining 'language and the actions into which it is woven' (what he calls the 'language-game'), a publicly available practice of using words: with pain, this involves crying, complaining, comforting, administering analgesics and so on (*Philosophical Investigations*, §7).

Another source of the idea of a private language is that, where thoughts and sensations are concerned, there is in the normal first-person case no gap between truthfulness and truth. Philosophers are inclined to interpret this as a special epistemic authority about one's own mind, whose contents are hidden from others. Wittgenstein's treatment of this has been one of the most disputed parts of his discussion; he has been interpreted as suggesting that the lack of first-person uncertainty in utterances like 'My leg hurts' derives from their being not genuine reports of inner occurrences but sophisticated groaning. It is now a short (and often-taken) step to infer that this involves a commitment to supposing that there is nothing behind the outer behaviour, and thus to a form of behaviourism (see Behaviourism, analytic). Such oscillation between privacy and behaviourism, as though the former were the sole alternative to the implausibility of the latter, is typical of the bind philosophers get into in this area, so that they cannot see that the blocking of this route to private language involves the attempt to undermine assumptions common to both. These include such ideas as that witnessing a person's acts is being confronted with neutral behavioural data from which mentality must be inferred or constructed, and that dealing with other human beings is always a matter of hypothesis rather than of attitude and response. But Wittgenstein's view is that these ideas misconstrue the metaphorical inner/outer contrast; his rejection of this construal is not a denial of the reality of the inner.

4 Kripke's Wittgenstein

Saul Kripke's account of Wittgenstein's treatment of rules and private language has generated much second-order discussion of a kind notably different from that of Wittgenstein himself. This is often carried on without apparent concern for the question of whether Wittgenstein's own arguments are captured by Kripke's interpretation. Thus the phrase 'private language argument' has acquired a further sense in which it refers to the arguments as given by Kripke.

Kripke's interpretation resembles that given here in its emphasis on the dependence of the discussion of private language on that of rule-following. But Kripke (1982: 68) says, 'The impossibility of private language emerges as a corollary of his sceptical solution of his own paradox', one stated at *Philosophical Investigations* §201: 'No course of action could be determined by a rule, because every course of action can be made out to accord with the rule.' Apparently overlooking Wittgenstein's reference to this paradox as involving a misunderstanding, Kripke takes it to be a profound sceptical problem about meaning. He formulates this problem in two ways.

The first formulation is this: there is no fact in which someone's meaning something consists. The absence of this fact, in Kripke's view, leads Wittgenstein to abandon the explanation of the meanings of statements like 'Smith

meant addition by 'plus'' in terms of truth-conditions; instead, they get explained in terms of assertibility-conditions, which involve actual (not merely potential) community agreement. (Hence the claim that this is a 'sceptical solution': Wittgenstein is supposed to concede to the sceptic the absence of truth-conditions for such statements.) This agreement, on Kripke's account, legitimizes the assertion that Smith meant addition by 'plus' despite there having been no fact of the matter, and it of course rules out the possibility of private language immediately (see Private states and language §4). This first formulation, though, relies on the assumption that we have some idea of what a fact is, independent of a statement's being true; and one of the fundamental lessons of *Philosophical Investigations* is that there is no such idea to be had, that the only route to the identification of facts is via the (often not easily discerned) uses of the expressions in which those facts are stated. These uses give us the truth-conditions.

The other formulation of the problem is that there is no nexus between someone's meaning something and their subsequent behaviour, so that, for example, my grasp of the rule governing the use of 'plus' does not determine that I shall produce a unique answer for each of indefinitely many new additions in the future. The impression that something is missing here, though, is a result of just that kind of confusion about determination identified in §1 above.

Kripke's account is of great intrinsic interest, but despite the acuteness of some of its observations, as an interpretation of the private language argument it is thus deficient, especially because of his unargued reliance on ideas which Wittgenstein himself argued against. This has not deterred philosophers from using it, as Norman Malcolm's earlier account was used, as a way of avoiding direct confrontation with Wittgenstein's own text.

See also: Consciousness; Criteria; Kripke, S.; Wittgenstein, L.

STEWART CANDLISH

References and further reading

Because of the great difficulty of the subject matter, none of this material is straightforward, but none is technical.

Boghossian, P.A. (1989) 'The rule-following considerations', *Mind* 98 (392): 507-49.(A critical survey of the literature concerning the Kripkean version of Wittgenstein's arguments.)

Candlish, S. (1980) 'The real private language argument', *Philosophy* 55 (211): 85-94.(A fuller version of some of the points made in §2 of this entry.)

Canfield, J.V. (ed.) (1986) *The Philosophy of Wittgenstein*, vol. 9, *The Private Language Argument*, New York: Garland.(An extensive collection of papers exhibiting the variety of interpretations of the argument.)

Canfield, J.V. (1986) *The Philosophy of Wittgenstein*, vol. 10, *Logical Necessity and Rules*, New York: Garland. (Contains some papers on the Kripke version.)

Canfield, J.V. (1991) 'Private language, *Philosophical Investigations* §258 and environs', in R. Arrington and H.-J. Glock (eds) *Wittgenstein's Philosophical Investigations: Text and Context*, London and New York: Routledge, 120-37.(A discussion of some common assumptions about how the argument works.)

Canfield, J.V. (1996) 'The Community View', *The Philosophical Review* 105 (4): 469-88.(A treatment of the issue of whether the private language argument requires language to be necessarily and actually, or merely potentially, social; examines both original texts and principal interpretations.)

Hacker, P.M.S. (1990) *Wittgenstein: Meaning and Mind*, *An Analytical Commentary on the Philosophical Investigations*, vol. 3, Oxford: Blackwell, 1-286. (Part of a massively detailed study of *Philosophical Investigations*; a good place to find out about textual sources, connections, ramifications, significance.)

Jones, O.R. (ed.) (1971) *The Private Language Argument*, London: Macmillan.(A collection of some of the earlier influential discussions of the argument. Also includes Wittgenstein 1968. Not much overlap with Canfield 1986.)

Kenny, A. (1973) *Wittgenstein*, London: Allen Lane: The Penguin Press, ch. 10.(One of the clearest accounts of the argument. Its version inspires, but differs from, that given in §2 above.)

Kripke, S. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell.(The final published version of Kripke's account.)

Malcolm, N. (1954) 'Wittgenstein's Philosophical Investigations', *The Philosophical Review* LXIII (4): 530-59. (Presents a highly influential early account of the private language argument.)

Winch, P. (1983) 'Facts and superfacts', *The Philosophical Quarterly* 33 (133): 398-404, repr. (slightly altered) in P. Winch *Trying to Make Sense*, Oxford: Blackwell, 54-63.(A clear and concise criticism of Kripke, from which some of the points in §4 above have been condensed.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell, 3rd edn 1967.(The seminal text. Spare and difficult without being obscure or technical. The central material is found in §§243-315, but the earlier discussions of ostension in §§1-36 and rule-following, §§142-242, are also drawn upon in the argument. See also §§316-97, and pages 220-4.)

Wittgenstein, L. (1968) 'Notes for lectures on "private experience" and "sense-data"', *The Philosophical Review* LXXVII: 275-320.(Also included in Jones 1971.)

Wittgenstein, L. (1975) *Philosophical Remarks*, Oxford: Blackwell.(A possible target of the private language argument.)

Private states and language

Something is 'private' if it can be known to one person only. Many have held that perceptions and bodily sensations are in this sense private, being knowable only by the person who experiences them. (You may know, it is often said, that we both call the same things 'green'; but whether they really look the same to me as they do to you, you have no means of telling.) Regarding the relation between private states and language two main questions have arisen:

(1) Could there be a 'private language', that is, a language in which a person communicates to themselves, or records for their own use, information about their own private states - this language being in principle incomprehensible to others, who do not know the nature of the events it is used to record. This question is primarily associated with Ludwig Wittgenstein.

(2) Can the nature of our private states affect the meaning of expressions in the public language, that is, the language we use for communicating with each other? Or must everything that affects the meaning of expressions in the public language be something which is itself public, and knowable in principle by anyone? Michael Dummett has argued that we must accept the second of these alternatives, and that this has far-reaching consequences in logic and metaphysics.

1 Epistemic privacy

The epistemically private is defined as whatever can in principle be known to one person only. The words 'in principle' are intended to exclude all cases in which something is in fact known to only one person, though it could be known to more. The only items that have generally been taken to be epistemically private in this very demanding sense are conscious mental states, primarily perceptual states and bodily sensations of a conscious subject, on the plausible grounds that nothing that a person can do can possibly give them a view of another's conscious states, but only of their bodily states and behaviour.

However, these grounds for taking such states to be epistemically private cannot by themselves decide the question whether there actually are any epistemically private items in the sense defined. Privacy, as we have seen, is defined in terms of what cannot be known to another, and from the fact that another cannot *experience* my headache it does not immediately follow that they cannot *know* what it is like. That must also depend on our understanding of knowledge, in particular on whether we can allow that what cannot be directly experienced by someone can nevertheless in some other way be known by them.

Discussion of this complex question is beyond the scope of this entry, in which it can only be remarked that some views of the nature of knowledge are not wholly unfavourable to the idea that the contents of another's consciousness may sometimes be known (see Reliabilism; Other minds). Here it will be assumed that there are epistemically private states of the type commonly suggested.

We have spoken above of *two* questions concerning private states and language, and it is important to realize that, although related, they must be considered separately if clarity is to be achieved. To see this, the reader should consider the following. An impressive argument (discussed in §§2-3) for denying that private states can play any part in the semantics of the language that we speak with each other begins from the premise that the nature of one's private states is unknown to anyone else, and concludes that if it in any way affected the meanings of our words we would not know what any other person meant, and so could not speak this language with mutual understanding - which is then taken to be absurd. As it stands this argument clearly has no application whatever to the first question, the possibility of a private language. For a private language is understood to be one in which no other speaker is involved, so that the fact that one's private states are unknowable to others is entirely beside the point. To reach the analogous conclusion it would have to be said that private states are unknowable not just to others but to their owners as well, and this much stronger claim clearly requires further and altogether different argument. Conversely, we shall also see (§4) that it is possible to argue against the feasibility of a private language using a line of thought that may not apply to the question about the public language.

2 Acquisition and manifestation

Two arguments now frequently offered for the conclusion that private states can play no role in determining the meaning of any expressions of our (public) language are known as the acquisition and manifestation arguments. Each assumes that, in so far as there exists a language in which we communicate with each other, the expressions of that language must be used in the same sense or meaning by all of us. The acquisition argument then asks how this state of affairs could come about; the manifestation argument asks how we could tell that it obtained.

Suppose that there are expressions of this public language whose meaning depends on the nature of certain of the private states of competent speakers. (Plausible examples are colour words - it might well be thought that what I mean by 'red' depends at least in part on how certain things look to me - and words for sensations such as 'pain'.) Now consider the situation of language learners acquiring an understanding of one such expression. *Ex hypothesi* this means that they must come to use it in connection with private states of their own of the same kind as those already associated with it by competent speakers; but since these states, being private, are not accessible to the learners, how can they hit on the right meaning? Learning the language would call for an absurdly improbable series of lucky guesses.

The manifestation argument views the process from the other side: how are the teachers to tell that the learners have learnt their lesson correctly? How can the learners show that they have grasped ('manifest their grasp of') the right meaning? For to grasp the right meaning is to connect with the expression particular private states rather than others; and because they are private that is something which the teachers are in principle unable to check. We could never know, in the case of these expressions, what any other speaker means by them.

We should note that neither of these arguments makes appeal to verificationism (see Meaning and verification) - both trade on the idea of what can be *known*, rather than of what is *meaningful*. The importance of this is not just that they avoid inheriting verificationism's weaknesses; it is also that otherwise they could not, without deep suspicion of circularity, be used to lead to a conclusion not altogether unlike verificationism - as they in fact have (see §5).

3 Responses to these arguments

One response, hinted at in §1, would be to deny that we have any states which are in principle unknowable to others. In view of the history of such sceptical debates this seems unlikely to lead to anything decisive.

A possible reply to the acquisition argument is as follows. Perhaps human beings naturally assume that others, when in broadly similar circumstances, experience inner states similar to those which they themselves experience. If that is so, and if the assumption is in most cases correct, then the problem posed by the acquisition argument can be answered. For since on this hypothesis the learner's private states resemble those of the teacher (when, for example, both are in the presence of the kind of object we call 'red') and are believed by both parties to do so, there is no difficulty in seeing how the learner can come to associate with the word 'red' private visual states which are of the same kind as those which the teacher associates with it.

But this does not settle the issue. For one thing, it makes no reply to the manifestation argument. For even if successful learning has occurred, how is anyone to tell? The teacher and the erstwhile learner have no way of knowing that they now mean the same. And if, as is often said, understanding someone is knowing what they mean, then they don't understand each other - contrary to the hypothesis that they are speaking a mutually intelligible language.

It is not obvious, however, that understanding someone does require knowledge of their meaning, if knowledge is taken to be more than confident true belief, as it nearly always is (see Knowledge, concept of §2). For if your words express a certain thought, and I rightly and confidently believe that that is what they express, and so respond appropriately, it seems that everything we want of understanding has been achieved, and the additional demand for knowledge begins to look arbitrary.

Still the matter is not settled, for it may be replied that even if knowledge is not a necessary condition of understanding, nonetheless we *do* know what others mean by their words; therefore any theory which allows private states an essential role in semantics must be wrong, since it is committed to denying that we have such knowledge.

That commitment cannot be disowned, since it follows immediately from the definition of privacy, but it may be

less damaging than it at first looks. Perhaps the claim to knowledge in this case, in so far as it implies the capacity to give good reasons for what we claim to know, is indeed dubious. Such a view is controversial, but certainly not obviously untenable. The idea that some of our most basic beliefs are produced by a psychological mechanism which is non-rational, in the sense that its operations are quite independent of any capacity we may have to *give reasons* for the beliefs it produces, has a reputable history (see for instance Hume, D.). To enable it to act successfully any animal, including ourselves and our ancestors, needs a reliable and rapid way of forming beliefs about its environment; so our basic belief-forming mechanisms cannot rest on reasoning, which is too slow and uncertain. And if a belief was not formed by reasoning there is little ground for thinking that it must nonetheless be certifiable, retrospectively, by rational argument, since it is not clear what additional practical benefit that would confer.

It is arguable, therefore, that the appeal to the threat of scepticism yields no decisive proof of the inadmissibility of private states in the semantics of public language.

4 The approach via rule-following

The question may also be viewed from another direction. Recent discussion of Wittgenstein's writings about rules (see Meaning and rule-following) has lead some (eminently Saul Kripke 1982) to suggest that the notion that two items are of the same kind can only have meaning against the background of a consensus in usage. This principle gives rise to the following argument.

Signs have meaning only because they are used in accordance with rules. So if anything plays an essential role in determining the meaning of a sign it must be capable of playing this role consistently, that is, in the same way on the various occasions on which the sign is used. Therefore private states can play an essential role only if it can be said that the same kind of private state was experienced on different occasions. But that, according to the above principle, can be so only if there is a communal practice with regard to these items among a number of speakers. And there can be no such communal practice - precisely because they are private. It follows that they play no part in the determination of meaning.

The 'consensus-principle', once accepted, instantly defeats the possibility of a private language, if that be understood as a language which of necessity only one person can speak. But that does not mean that it immediately dismisses private states from any position in the semantics of public languages. For if one can have true beliefs about another person's private states (as perhaps one can, so long as the hypothesis advanced at the beginning of §3 remains unrefuted) then there can be such a thing as agreement about these states, and even disagreement, at least where insincerity is suspected. And then if this is held not to satisfy the demand for communal practice, it needs to be explained why the consensus required should need to be based on mutual knowledge rather than just confident belief - no easy task, since it has not proved easy convincingly to establish *any* form of the consensus principle.

5 Further consequences

This question has not normally been pursued for its own sake, but rather for that of its alleged consequences. M.A.E. Dummett, having concluded that a speaker's grasp of the meaning of a sentence must consist in their potential for outward, publicly accessible behaviour, tentatively suggests a further conclusion: it must consist in the capacity to recognize whatever state of affairs it is that makes the sentence true, in other words to tell that it is true, if and when it is. We can grasp no meaning, and so have no grasp of truth, beyond what we can in principle recognize as true; and it is this which is held to define antirealism and require the abandonment, characteristic of intuitionist mathematics, of the law of excluded middle as a generally valid logical principle.

The response to this train of thought must depend on the answers to a number of questions. Is it clear that the recognition of the fact that the truth conditions of a sentence obtain is the only (publicly accessible) way to demonstrate ('manifest') understanding of it? Is it even clear that the act of recognizing that certain conditions obtain *is* in the required sense publicly accessible? This cannot just be assumed. Recognition is an intensional state, since one can recognize the fact that p without recognizing the fact that q, even if these two facts always obtain together; we are therefore lead into the difficulties connected with such states, in particular the question whether they can be captured in purely physical behaviour (see Intensionality;Intentionality §2).

But neither should it be assumed that problems about meaning and understanding are easily resolved, or Dummett's conclusions easily avoided, if we are allowed the contrary assumption, that private states may be given a role in semantics. For that it is still too unclear what kinds of private state there are, and what their role could be, and hence what difference it makes if we admit them.

See also: Intuitionistic logic and antirealism; Private language argument

EDWARD CRAIG

References and further reading

All these items involve intricate argument but little or no technicality.

Craig, E.J. (1997) 'Meaning and Privacy', in *A Companion to the Philosophy of Language*, Oxford: Blackwell, 127-45.(Expansion of the material in this entry; also contains a section on the private language argument.)

Dummett, M.A.E. (1973) 'The Philosophical Basis of Intuitionistic Logic', in *Truth and Other Enigmas*, London: Duckworth, 1978, esp. 215-27.(Presents the arguments of §2 and suggests their connection with antirealism as discussed in §5.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell.(Expounds the considerations leading to the 'consensus' principle as mentioned in §4. A particularly fine piece of writing.)

Wright, C.J.G. (1987) 'Introduction', to *Realism, Meaning and Truth*, Oxford: Blackwell, esp. 13-23.(Expounds and discusses the arguments from acquisition and manifestation and connects them with antirealism.)

Proper names

The Roman general Julius Caesar was assassinated on 14 March 44 BC by conspirators led by Brutus and Cassius. It is a remarkable fact that, in so informing or reminding the reader, the proper names 'Julius Caesar', 'Brutus' and 'Cassius' are used to refer to three people each of whom has been dead for about two thousand years. Our eyes could not be used to see any of them, nor our voices to talk to them, yet we can refer to them with our words.

The central philosophical issue about proper names is how this sort of thing is possible: what exactly is the mechanism by which the user of a name succeeds in referring with the name to its bearer? As the example indicates, whatever the mechanism is, it must be something that can relate the use of a name to its bearer even after the bearer has ceased to exist.

In modern philosophy of language there are two main views about the nature of the mechanism. On one account, which originated with Frege, a use of a name expresses a conception or way of thinking of an object, and the name refers to whatever object fits, or best fits, that conception or way of thinking. Thus with 'Cassius', for example, I may associate the conception 'the conspirator whom Caesar suspected because of his size' (recalling a famous speech in Shakespeare's Julius Caesar). Conception theories are usually called 'sense' theories, after Frege's term 'Sinn'. The other account is the 'historical chain' theory, due to Kripke and Geach. In Geach's words, ' for the use of a word as a proper name there must in the first instance e someone acquainted with the object named…. But…the use of a given name for a given object…can be handed on from one generation to another…Plato knew Socrates, and Aristotle knew Plato, and Theophrastus knew Aristotle, and so on in apostolic succession down to our own times. That is why we can legitimately use "Socrates" as a name the way we do' (1969-70: 288-9).

1 Sense theories: Introduction

The idea that names express reference-determining conceptions or 'senses' originated with Frege (1892). But Frege was led to it not by wondering about the mechanism of reference, but rather about the difference in 'cognitive value' between two true identity statements of the respective forms $a = a$ and $a = b$ (for example, 'Clark Kent is Clark Kent' and 'Clark Kent is Superman'). Statements of the former sort, he says, are a priori, while the latter 'often contain very valuable extensions of our knowledge and cannot always be established a priori' ([1892] 1980: 56). Frege takes this to show that the propositions expressed by $a = a$ and $a = b$ cannot be the same even when *a* and *b* are names of the same person. Yet if the names do corefer, there is no difference in the references of the constituents of $a = a$ and $a = b$; moreover, these two sentences are assembled in the same way. So either they cannot express different propositions, or else - and this is the inference Frege drew - what determines the proposition a sentence *S* expresses cannot just have to do with the structure of *S* and the *references* of its constituent words and phrases (see Frege, G. §§3-4; Sense and reference §1).

Frege's proposal was that in addition to possessing a reference, any meaningful expression has a sense, the sense being a 'way of thinking' of the reference. The proposition a sentence expresses is determined by the senses of the words in it, not their references. Thus 'Clark Kent is Clark Kent' and 'Clark Kent is Superman' express different propositions because 'Clark Kent' and 'Superman' express different senses, senses which happen to be ways of thinking of the same reference. Taking the Superman fiction to be fact, the sense of 'Clark Kent' might be 'the mild-mannered reporter on *The Daily Planet* who has a crush on Lois Lane' while the sense of 'Superman' might be 'the blue-suited extraterrestrial who flies'. Here we specify different conceptions of the same individual in two definite descriptions (expressions of the form 'the so-and-so' - see Descriptions). The proposition that Superman is Clark Kent therefore has the content that the blue-suited extraterrestrial who flies is the mild-mannered reporter on *The Daily Planet* who has a crush on Lois Lane, which may indeed be, in Frege's phrase, an extension of our knowledge.

It was Frege's view that in an ideal language each name would have a fixed sense and reference for everyone; in an ordinary natural language, there are names that fail to refer, and users may fail to agree on the sense of a name for a specific individual, which Frege thought of as deficiencies of natural language. In a modification of Frege's views, Searle (1958) allowed that a name may be associated with a whole range of descriptions, different users using different ranges; the bearer of the name need not satisfy all the conditions mentioned in the descriptions, only 'a sufficient number', an intentionally vague condition.

Sense theories along such lines as these have been called 'famous deeds' sense theories. Famous deeds sense theories appear to have been conclusively refuted by Kripke (1972, 1980). Kripke objects to such sense theories both as they respond to Frege's own puzzle about the difference between $a = a$ and $a = b$, and as they address our initial question about the mechanism of reference. Let us consider these two issues in turn.

2 Sense theories: Do names express senses?

According to Kripke, the propositions one expresses using names do not involve senses or ways of thinking expressed by the names, otherwise some such propositions would be both metaphysically necessary and a priori, which they are clearly not. For example, suppose the sense of 'Aristotle' is 'the pupil of Plato who tutored Alexander'. Then the sentence

(1) Aristotle was a pupil of Plato

would express the proposition with the content

(2) The pupil of Plato who tutored Alexander was a pupil of Plato.

But (2) is, in a certain sense, necessary, while (1) is not. There is no way things could have gone in which (a) a unique pupil of Plato who tutored Alexander exists and (b) that person was not a pupil of Plato. On the other hand, there are many ways things could have gone in which (a′) Aristotle exists but (b′) Aristotle was not a pupil of Plato (for example, he died young). In other words, granted that there is such a person as Aristotle, he may be a pupil of Plato or he may not be. But granted that there is such a person as the pupil of Plato who tutored Alexander, it follows that he is a pupil of Plato.

This example brings out Kripke's famous distinction between 'rigid' and 'non-rigid' designators. In thinking about or describing other ways things could have gone (other possible worlds) we use a proper name such as 'Aristotle' consistently to denote the same person; this makes proper names rigid designators. But we use definite descriptions such as 'the pupil of Plato who tutored Alexander' differently. With respect to the actual world, this description picks out Aristotle; with respect to a possible world where someone else is the one and only pupil of Plato who tutored Alexander, the description picks out that other person, not Aristotle; and with respect to a possible world where either Plato or Alexander does not exist, the description fails to pick out anyone, even if Aristotle does exist. Hence such descriptions are non-rigid designators. Certain descriptions, such as 'the positive square root of 9', are as rigid as proper names, but this is on account of their subject matter, not their semantic role. It is because the typical famous deeds description is non-rigid that the contrast between pairs such as (1) and (2) *vis-à-vis* necessity and contingency arises (see Reference §2).

Just as (1) and (2) differ in modal status, they differ epistemically: bracketing the question of existence, (2) is, in a limited sense, knowable a priori, while (1) is not. That is, granted that there was such a person as Aristotle, it is a further, empirical question whether he was a pupil of Plato (there may be a controversy among historians about this). But granted that there was such a person as the pupil of Plato who tutored Alexander, it is not a further question, *a fortiori* not an empirical one, whether he was a pupil of Plato.

The same objections arise to Searle's modified version of the sense theory. For example, no matter what range of descriptions we associate with a name, they will generate statements that are a priori in the manner of (2). If being ϕ logically implies being ϕ', then

(3) The thing which is *F*-and-*G* or *G*-and-*H* is *F′* or *G′* or *H′*

is essentially the same as (2), just more complicated. But if the predicates *F*, *G* and *H* encapsulate famous deeds, the corresponding

(4) *NN* is *F′* or *G′* or *H′*,

where '*NN*' is the name with which 'the thing which is *F* or *G* or *H*' is associated, will be no more a priori than (1) (Kripke 1980).

3 Sense theories: Do senses determine reference?

Kripke (1972, 1980) demonstrates another flaw in famous deeds sense theories, namely, that they do not provide an adequate answer to the question about the mechanism of reference. This is because (1) the likely candidate for the sense of a name may pick out an object which is not in fact the name's bearer, or may fail to pick out anything, and (2) we can succeed in referring with a name even when we do not have a 'famous deeds' description associated with it.

Kripke illustrates (1) with two examples. If any description is associated with the name 'Gödel' it is 'the discoverer of the incompleteness of arithmetic'. Does this mean that 'Gödel' refers to that person? What if the theorem was actually proved by Schmidt, who died in strange circumstances, and Gödel got hold of Schmidt's work and represented it as his own? The very fact that we can understand this 'what if' shows that the reference of 'Gödel' is not fixed as whoever discovered the incompleteness of arithmetic. And though this example is fictional, there are similar actual cases: Peano's Axioms are not due to Peano; and Einstein was not the inventor of the atomic bomb (Kripke 1980). For a case where there are descriptions that do not pick out any object although the relevant name still refers, Kripke gives the example of the prophet Jonah, who really existed (according to the scholarly consensus), but whose career as described in the Bible is essentially fictitious. Again, the mere intelligibility of the claim 'Jonah was a historical person but everything uniquely identifying that the Bible says about him is fictitious' is enough to show that the reference of 'Jonah' is not fixed as the Hebrew prophet who was swallowed by a whale, or by any other condition, no matter how complicated or disjunctive, derived from the Book of Jonah.

As for (2), successful reference without associated (definite) descriptions, Kripke points out that most people can use the names 'Richard Feynman' and 'Murray Gell-Mann' to refer to those two people, but that, at best, all the typical person knows about either is that he is a famous physicist who won a Nobel Prize (this was before Feynman achieved popular fame for his role in the *Challenger* disaster inquiry). So we have difference in reference with no difference in associated descriptions; therefore associated descriptions are not at the heart of how proper name reference works.

Perhaps these examples only establish such a conclusion for descriptions that encapsulate famous deeds. Kripke considers some other approaches and concludes that they violate an important non-circularity condition: that candidate descriptions must not themselves embed the notion of reference in a way that cannot eventually be eliminated. For example, we might suggest that the reference of 'Socrates' is fixed as 'the man called "Socrates"'. Since 'called' just means 'referred to as', we do not explain how reference to Socrates is possible in this way: what we want to know is how Socrates *gets to be* the man referred to as 'Socrates'. So this goes nowhere as a proposal about the mechanism of reference, and Kripke plausibly argues that the same would be true for more complicated versions of the idea, for example, that 'Gödel' refers to the person to whom the proof of the incompleteness of arithmetic is commonly attributed. So these attempts at a non-famous-deeds description theory fail. However, in view of Kripke's critique of the famous deeds approach, it seems that if any sense theory is to work, it will have to be a non-famous-deeds account of some sort.

4 Historical chains

The main competing account of the mechanism of reference is the Geach-Kripke historical chain account, in which competence to refer with the name is transmitted across generations in the style adverted to in the quotation from Geach in this entry's prologue (1969-70). The historical chain account is sometimes called the 'causal' chain account, since it is held that the links in the chain are forged by transactions of a causal sort. In Kripke's own version, which he says is a 'picture' rather than a theory (1980: 97), a name is introduced into a community by some 'initial baptism' of an object with the name, and then the name is passed on from link to link, it being required at each step that the receiver of the name 'intend…to use it with the same reference as the [person] from whom he heard it' (1980: 96).

If we regard being told about, or otherwise hearing about, an object, as a way of becoming 'acquainted' with it, then the Geach-Kripke account instantiates an approach championed by Russell (1918), who made acquaintance with an object necessary for referring to it. However, Russell's notion of acquaintance was rather idiosyncratic: apart from my own sense-data, my self, universals and perhaps the present moment, I lack acquaintance with things, according to Russell. Since the ordinary names I use, ostensibly for other people and things, are not names of sense-data, Russell claimed that in a 'logical' sense, ordinary names are not 'proper' names. Rather, he

suggested, they are definite descriptions in disguise (see Russell, B.A.W. §9). For this reason, Kripke sometimes calls the view he opposes the 'Frege-Russell theory of names'. The point to bear in mind is that Frege and Russell had different accounts of how reference works (so there is no 'Frege-Russell theory of reference'), but because on Russell's account ordinary names do not really refer, the two philosophers end up saying similar-sounding things about such names.

Returning to Kripke's account of the mechanism of reference in terms of informational exchanges in which the intention to preserve reference is present, the obvious question is whether it is any improvement on circular description theories. After all, the notion of reference enters explicitly into Kripke's necessary condition for successful passing on of the name, and it also enters at the start of the chain, where some kind of demonstrative reference to the object being baptized is standardly made. To put the same question another way, if this account is explanatory, would there be anything wrong with a description theory which attributed to a name '*NN*' and a user of the name *U* the sense 'the object at the start of the chain of reference-preserving links by which I came into mastery of this name', in which 'I' refers to *U*?

One problem with this version of a description theory is that no ordinary speaker of a natural language associates any such description with a proper name, since the description embodies a philosophical theory, and it is difficult to see how one could justify claiming that the association is 'implicit'. Here there is a contrast with Kripke's proposal, which only requires speakers to have the intention to preserve the reference of the new name they have just learned, and it is surely plausible that speakers do have such an intention. Still, there is some sense in which the fundamental nature of the mechanism of reference is left unexplained. For instance, what exactly is it about an actual baptismal service, or about parents announcing their choice of name, that causes a new 'common currency' (see Kaplan 1990) name to be added to the language? Or is the request for further explanation here a demand for a 'reductive' account of reference in terms of non-semantic notions (see Reference §8), something perhaps impossible?

The causal chain picture is not without other difficulties, as Kripke mentions. For instance, straightforward application of the picture could lead to the conclusion that 'Santa Claus' is the name of a certain central European king, which seems wrong. Evans (1973) proposed a different causal account on which the 'causal source' of the information associated with the name determines whom it refers to, though in a possibly complex way. Evans observes that the reference of a name in a linguistic community can change over time if it was originally a name for *x* that was mistakenly but consistently misapplied to *y: y* would ultimately become the bearer of the name, being the dominant causal source of the information associated with it (unless, as Evans notes, younger members of the community defer in their use of the name towards those for whom it was once a name for *x*). Evans subsequently elaborated his account (1982: ch. 11).

5 Direct reference and Frege's puzzle

A striking feature of the Geach-Kripke picture is that it leaves no role for sense to play as that which determines reference. The only candidate for the 'meaning' of a name is therefore the name's reference itself. A theory which claims that the meaning of a name is just its reference, pure and simple, is often called a 'direct reference' theory, 'direct' signifying that reference is not mediated via sense. Such theories trace their origins through Russell back to Mill (1843); more recently, the idea is prominent in the writings of Marcus on reference (1961), and has subsequently been developed and defended in a sustained form by Salmon (1986) and Soames (1994).

The main problem which direct reference theories face is the puzzle about the difference between $a = a$ and $a = b$. The general 'failure of substitutivity' puzzle is the puzzle of how it is possible for the meanings of two sentences to differ if the sentences have the same structure and their corresponding parts have the same meaning. For example,

(5) It is self-evident to any rational thinker that if Superman exists, then Superman = Superman

seems true and 'Superman' and 'Clark Kent' have the same reference. Therefore, according to direct reference theory, (5) should have the same meaning as

(6) It is self-evident to any rational thinker that if Superman exists, then Superman = Clark Kent.

But at first sight, (6) does not even have the same truth-value as (5), never mind the same meaning.

Failure of substitutivity is handled straightforwardly on a sense theory of names which allows that coreferential names can have different senses. The claim would be that (not merely the reference but) the sense of 'Superman' enters into the truth-condition of (5), while in (6) the sense of 'Clark Kent' is also involved. There are various mechanisms which might be invoked here, Frege's own being the simplest: in (5) and (6), according to Frege, the proper names refer to their *senses*, rather than to the person they normally refer to (see Propositional attitude statements §1; Sense and reference §5). But the trouble with any explanation of substitutivity failure for proper names that invokes senses is that sense theories of proper names have been so thoroughly battered by Kripke.

Direct reference theory seems to have less to work with to explain the semantic difference between (5) and (6). On perhaps the best known version of the theory, Salmon's, it is simply denied that there is a semantic difference between (5) and (6). The appearance that substitution changes truth-value in cases such as (5) and (6) is explained as being due to pragmatic effects (see Pragmatics). Apparently, the only alternative for a direct reference theorist is to identify some assumption in the argument that substitution of coreferring names ought not to change meaning and deny that assumption. This strategy produces a possible target. We suggested that what makes substitutivity failure puzzling is that sentences with the same structure whose corresponding parts have the same meaning should themselves have the same meaning. But this presupposes that the meaning of the entire sentence is wholly determined by the meaning of its parts and their manner of composition (see Compositionality). However, the notion of structure alluded to in the phrase 'manner of composition' is one which is sensitive only to syntactic categories of expression, not to the identity of expressions. There is another, more 'logical' notion of structure, on which use of a different word, even one with the same meaning, can disrupt structure (Putnam 1954). In this logical sense, the structure of 'if Superman exists, then Superman = Superman' is 'if Et then $t\ t$', while 'if Superman exists, then Superman = Clark Kent' has the different structure 'if Et then $t = t$Čast;'. If we hold that substitution is acceptable only when it does not change logical structure, then from (5) we can infer merely

(7) It is self-evident to any rational thinker that if Clark Kent exists, then Clark Kent = Clark Kent

which is presumably true if (5) is. Of course, this does not help with cases of substitutivity failure in which there is only one occurrence of the name being substituted ('Lois believes that Superman is an extraterrestrial'), but some direct reference theorists have tried to develop a more general notion of structure for these cases (Taschek 1995).

It is the handling of substitutivity puzzles that is the deciding issue between sense theories and direct reference theories of proper names. Sense theorists need to find a viable account of the senses of names and direct reference theorists need to find a persuasive account either of why substitutivity fails in cases such as (5), or else of why it gives such a convincing appearance of doing so.

See also: De re/de dicto; Kripke, S.A.; Logical and mathematical terms, glossary of

GRAEME FORBES

References and further reading

Burge, T. (1973) 'Reference and Proper Names', *Journal of Philosophy* 70: 425-39. (Argues that, semantically, proper names should be understood as certain kinds of predicate.)

Deutsch, H. (1989) 'On Direct Reference', in J. Almog, J. Perry and H.K. Wettstein (eds) *Themes from Kaplan*, Oxford and New York: Oxford University Press, 167-95. (Applies a Kaplan-style approach to context-dependence to proper names.)

Devitt, M. (1981) *Designation*, New York: Columbia University Press.(Develops a causal theory of reference using the 'language of thought' hypothesis.)

Evans, G. (1973) 'The Causal Theory of Names', *Proceedings of the Aristotelian Society*, supplementary vol. 47: 187-208; repr. in *Collected Papers*, Oxford: Oxford University Press, 1985, 1-24.(A very accessible and insightful discussion of Kripke's account.)

Evans, G. (1979) 'Reference and Contingency', *The Monist* 62: 161-84; repr. in *Collected Papers*, Oxford: Oxford University Press, 1985, 178-213.(Discusses a category of proper name not considered in this article, which Evans calls 'descriptive' names. These are names whose reference is explicitly fixed by a definite description, and which figure in Kripke-style examples of the contingent a priori.)

Evans, G. (1982) *The Varieties of Reference*, Oxford and New York: Oxford University Press.(Published after

Evans' death; some parts are unfinished. As well as Evans' final thoughts on proper names, includes influential discussion on Frege, Russell and the notion of an 'object-dependent' thought.)

Forbes, G. (1990) 'The Indispensability of *Sinn*', *Philosophical Review* 99: 535-63. (Develops a non-famous-deeds sense theory of names and applies it to substitutivity failure.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. M. Black, 'On Sense and Reference', in *Translations from the Philosophical Writings of Gottlob Frege*, ed. P.T. Geach and M. Black, Oxford and New York: Blackwell, 3rd edn, 1980.(Widely regarded as the paper which inaugurated modern philosophy of language.)

Geach, P.T. (1969-70) 'The Perils of Pauline', *Review of Metaphysics* 23: 287-300; repr. in *Logic Matters*, Berkeley, CA: University of California Press, 1972, 153-65. (States the historical chain view and discusses the use of names in intentional contexts.)

Jubien, M. (1993) *Ontology and the Fallacy of Reference*, Cambridge and New York: Cambridge University Press.(An unusual but interesting account of reference, based on 'revisionist' metaphysics.)

Kaplan, D. (1990) 'Words', *Proceedings of the Aristotelian Society*, supplementary vol. 64: 93-117.(Includes an original and accessible discussion of how names should be individuated.)

Kripke, S.A. (1972) 'Naming and Necessity', in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 252-355.(One of the most influential and widely discussed works in post-war philosophy.)

Kripke, S.A. (1979) 'A Puzzle About Belief', in A. Margalit (ed.) *Meaning and Use*, Dordrecht: Reidel. (Develops two famous cases - the London/Londres case and the Paderewski case - in defence of the thesis that substitution puzzles do not necessarily tell against the principle of substitutivity.)

Kripke, S.A. (1980) *Naming and Necessity*, Oxford and New York: Blackwell.(The book version of Kripke (1972).)

Marcus, R.B. (1961) 'Modalities and Intensional Languages', *Synthèse* 13: 303-22. (Includes an early statement of contemporary direct reference theory.)

Mill, J.S. (1843) *A System of Logic*, London: Parker; repr. in *The Collected Works of John Stuart Mill*, ed. J.M. Robson, London: Routledge, 1996, vol. 7.(Draws a famous distinction between the denotation and the connotation of a word and defends a metalinguistic analysis of identity sentences.)

Putnam, H. (1954) 'Synonymy and the Analysis of Belief Sentences', *Analysis* 14: 114-22.(Attacks compositionality assumptions in substitution puzzles.)

Recanati, F. (1993) *Direct Reference*, Oxford and New York: Blackwell. (Distinguishes a sense in which names are directly referential from two other semantically relevant senses in which they express modes of presentation. Also includes a good discussion of how 'what is said' by an utterance is determined.)

Russell, B.A.W. (1918) 'The Philosophy of Logical Atomism', *The Monist* 28: 495-527; 29: 32-63, 190-222, 345-80; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992.(One of many possible sources of Russell's views, which changed throughout his long life. Lectures II and VI are the most germane to this entry.)

Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.(The most complete attempt by a direct reference theorist to account for the problem of (apparent) failure of substitutivity.)

Searle, J. (1958) 'Proper Names', *Mind* 67: 166-73.(A significant elaboration of broadly Fregean themes.)

Soames, S. (1994) 'Attitudes and Anaphora', in J.E. Tomberlin (ed.) *Philosophical Perspectives: 8 Logic and Language*, Atascadero, CA: Ridgeview, 251-72.(Within the context of direct reference theory, investigates the behaviour of anaphoric pronouns in propositional attitude ascriptions. Mostly nontechnical, but rigorous.)

Strawson, P.F. (1959) *Individuals*, London: Methuen.(Chapter 6 includes an elaborated Fregean account that Kripke often uses as a stalking horse.)

Taschek, W. (1995) 'Belief, Substitution, and Logical Structure', *Noûs* 29: 71-95. (Extends the approach of Putnam (1954) to account for a wider range of substitution puzzles.)

Propositional attitude statements

Propositional attitude statements - statements about our beliefs, desires, hopes and fears - exhibit certain logical peculiarities. For example, in apparent violation of Leibniz's law of the indiscernibility of identicals, we cannot freely substitute expressions which designate the same object within such statements. According to Leibniz's law, every instance of the following scheme is valid:

a = b
*F(*a*)*
*Therefore, F(*b*)*

The validity of Leibniz's law seems beyond question. It says, in effect, that if an object has a certain property, then anything identical to that object also has that property. Valid instances abound. But consider the following apparently invalid instance:

(1) Hesperus is Phosphorus
(2) Hammurabi believed that Hesperus often rose in the evening
(3) Therefore, Hammurabi believed that 'Phosphorus' often rose in the evening.

If we take 'Hammurabi believed that…often rose in the evening' to serve as the predicate F and 'Hesperus' and 'Phosphorus' to be a *and* b *respectively, this argument appears to be an instance of Leibniz's law. Yet (3) apparently fails to follow from (1) and (2). Hammurabi believed that Hesperus and Phosphorus were two heavenly bodies not one. And he believed that Hesperus did, but that Phosphorus did not rise in the evening.*

We have derived a false conclusion from true premises and an apparently valid law. If that law is really valid, then our argument had better not be a genuine instance of the law. The tempting conclusion, widely accepted, is that we were wrong to construe propositional attitude statements as simple predications. We should not, that is, construe 'Hammurabi believed that…often rose in the evening' to be just a long predicate with the semantic function of attributing some property to the object commonly denoted by 'Hesperus' and 'Phosphorus'. But then the question arises: if attitude reports are not simple predications, what are they? Philosophers have disagreed sharply in their answers. Moreover, their disagreements are intimately connected to a wide range of deep issues about the nature of meaning and reference.

1 Frege on attitude statements

Gottlob Frege was among the first to offer systematic explanations of the form and content of propositional attitude statements (Frege 1892). Central to Frege's treatment of attitude statements is his distinction between sense and reference (see Sense and reference). Frege held that each significant linguistic expression plays two distinct but related semantic roles: it 'denotes' a reference and 'expresses' a sense. For example, the reference of a singular term like a proper name is typically some individual object. A sense, on the other hand, is a way of being given a reference. The sense of a proper name, for example, contains a condition, the unique satisfaction of which by an object *o* is necessary and sufficient to determine *o* as its reference. Perhaps the following captures the sense of 'Phosphorus':

o is Phosphorus just in case *o* is the last celestial object visible in the early morning sky just before sunrise.

The distinction between sense and reference holds not just for proper names but also for sentences. Frege took the reference of a complete declarative sentence to be its truth-value. A declarative sentence is thus a name of sorts - a name for a truth-value. So, for example, the reference of:

(4) Phosphorus can often be seen in the evening sky

is the True, as Frege called it. On the other hand, the sense of a complete declarative sentence is supposed to be a thought, according to Frege. The sense of (4) is the thought that Phosphorus can often be seen in the evening sky.

Frege held that that two expressions may denote the same reference while expressing distinct senses. 'Hesperus' and 'Phosphorus' denote the same object, but differ in sense, from Frege's viewpoint. Something like the

following satisfaction condition arguably captures the sense of 'Hesperus':

o is Hesperus just in case *o* is the first celestial object visible in the early evening sky after sunset.

Similarly, although sentence (4) and

(5) Hesperus can often be seen in the evening sky

denote the same reference - namely, the True - they express different thoughts. For the thought that Phosphorus can often be seen in the evening sky is, according to Frege, distinct from the thought that Hesperus can often be seen in the evening sky. One indicator of their distinctness is the possibility that a rational cognizer can simultaneously believe the one, while disbelieving the other. Hammurabi was just such a cognizer.

It is worth enumerating three essential characteristics of Fregean thoughts. First, thoughts are the primary bearers of truth-conditions. The very identity of a thought is constituted by its having the truth-conditions that it has. Sentences, by contrast, have their truth-conditions only derivatively, by being associated with a thought, in virtue of the conventions of a language. Second, thoughts are composite structured entities. A thought is a whole, composed of constituent parts. The parts of thoughts are themselves senses. For example, the thought that Hesperus is often visible in the evening has the sense of 'Hesperus' as a constituent part. Third, thoughts are the objects of the propositional attitudes. To believe that Hesperus is often visible in the evening is to stand in a certain relation - the believing relation - to the thought that Hesperus is often visible in the evening. To wonder whether Hesperus is often visible in the evening is to stand in a different relation - the wondering relation - to that same thought.

Frege endorses a principle of compositionality for both sense and reference (see Compositionality). He believes that both the reference and the sense of a complex expression are a function of the references and senses, respectively, of the constituent parts of that sentence (and the way those parts combine to form a whole). Two sentences which differ only by constituents which themselves have the same reference, should have the same truth-value. Similarly, two sentences which differ only by constituents which express the same sense, should themselves express the same thought. Now Frege is convinced that these compositionality principles are nowhere violated. But then the original quandary about attitude sentences remains.

Frege's way out is to say that despite our initial assessment 'Hesperus' as it occurs in (2) and 'Phosphorus' as it occurs in (3) do not share a reference. Call the sense and reference of a term as it occurs in direct discourse its 'customary' sense and reference. Frege claims that when an expression is embedded within a that-clause, it undergoes a shift in both its sense and its reference. In particular, he holds that obliquely occurring expressions, as he called them, denote not their customary referents, but their customary senses. Moreover, an obliquely occurring term denotes a new sense: what he calls its 'oblique' or 'indirect' sense. An indirect sense is a mode of presentation not of a term's customary reference, but of its customary sense. Thus in (2) 'Hesperus' refers not to the planet Venus but to the customary sense of 'Hesperus'. Similarly in (3), 'Phosphorus' denotes the customary sense of 'Phosphorus'. Indeed, a whole sentence undergoes a shift in sense and reference when it is joined with 'that' to form a clause. The that-clause in (2), for example, does not denote a truth-value, but the thought customarily expressed by 'Hesperus often rose in the evening' when it stands on its own. It denotes, that is, the thought that Hesperus often rose in the evening. And (2) as a whole says that Hammurabi stands in the believing relation to that thought. Similarly, the that-clause in (3) denotes the thought that Phosphorus often rose in the evening. And (3) attributes to Hammurabi the property of believing that thought.

Clearly, if Frege is correct, the invalidity of our original argument does not, after all, violate Leibniz's law. Since the customary sense of 'Hesperus' is not identical to the customary sense of 'Phosphorus' our original argument turns out not to involve the substitution of co-referring terms.

2 Referential opacity

Philosophers have sometimes questioned the coherence of Fregean senses. W.V. Quine, for example, has argued that senses and related intensional notions are 'creatures of darkness' (Quine 1956). There is no saying, according to Quine, for arbitrary senses *x* and *y*, when *x* and *y* are the same sense and when they are two distinct senses (see Quine, W.V. §8; Radical translation and radical interpretation §§2-3). There can be no entities, he insists, where

there are no determinate criteria of identity. But if senses go, so too must Frege's account of propositional attitude statements.

Quine endorses Frege's negative conclusion that obliquely occurring terms do not play their customary referential roles. He offers, however, a markedly different diagnosis of that failure. Like Frege, Quine believes that where reference happens substitutivity reigns - that is, Leibniz's law holds. But unlike Frege, Quine takes the failure of substitutivity within attitude statements at face value. And he concludes, in effect, that reference cannot be happening within the context of propositional attitude statements or that-clauses generally. That-clauses, he claims, are 'referentially opaque'. Embed a term which elsewhere functions referentially within such a clause and substitution is no longer permissible, not because reference shifts, but because the embedded term is stripped of its referential function. Quine suggests, in fact, that it is a mistake to speak of 'embedding' a referring expression within a that-clause at all. He suggests that an expression like 'believed that Hesperus often rose in the evening' functions rather like a primitive one-place predicate with no 'logically germane' constituent parts. In that case, sentences like (2) are misleadingly spelled. A more revealing spelling would treat this predicate as a single, though very long word somewhat like: 'believed-that-Hesperus-often-rose-in-the-evening'. The presence of the string 'Hesperus' in this very long word is merely an accident of orthography. It would be no more correct to regard 'Hesperus' as a grammatical constituent of this very long word, than it would be to regard the word 'cat' as grammatical constituent of the word 'cattle'.

Referential opacity explains another logical peculiarity of that-clauses, according to Quine. Consider the law of existential generalization:

F(a)
Therefore, $(\exists x)F(x)$

Informally, this scheme says that if a has F then there is something or other which has F. Existential generalization allows us to infer from the premise that 'John loves Mary' to the conclusion that 'John loves someone or other'. When we try to apply this evidently valid scheme within the context of a propositional attitude verb, we are led into apparent contradiction. Consider the following sentence:

(6) The president suspects that someone is a spy.

As it stands, (6) is subject to two different readings. On one reading, (6) says that the president 'suspects true', as we might put it, a certain statement or proposition - the statement or proposition that spying sometimes happens. This reading has been variously called the '*de dicto*', 'notional', or 'referentially opaque' reading (see De re/de dicto). There is a second reading of (6) which attributes to the president not just the generalized suspicion that spying sometimes happens, but the state of being 'suspicious of' some person or other. We might paraphrase this reading of (6) by:

(6b) The president suspects someone of being a spy

or by:

(6c) There is someone whom the president suspects of being a spy.

This reading is variously called the '*de re*', 'relational', or 'referentially transparent' reading. The two readings are not generally equivalent. It is possible to suspect that spying happens even when one has no suspicions about any particular person. The ambiguity in sentences like (6) is sometimes traced to an ambiguity of quantifier scope. On the so-called *de dicto* or opaque reading of (6), the quantifier has narrow scope relative to 'believes'. On the so-called *de re* or transparent reading, the quantifier has wide scope relative to 'believes'. As a first pass, we might represent the *de dicto* reading by (6d) and the *de re* reading by (6e):

(6d) The president suspects that $(\exists x)(x$ is a spy$)$
(6e) $(\exists x)($The president suspects that x is a spy$)$.

Quine has argued that constructions like (6e), in which a quantifier sitting outside the scope of the attitude verb 'reaches in' to bind a variable that lies within the scope of that attitude verb, are deeply problematic. We cannot,

he claims, 'quantify into' attitude contexts (Quine 1956). Or to put it in slightly different terms, propositional attitude verbs appear to block the interior reach of exterior quantifiers.

Suppose that the president has observed Stanley P. Young, a certain low-level White House aide, behaving in a rather furtive manner. He says with great conviction to the chief of White House security, 'Stanley P. Young is a spy.' That seems a sufficient basis for attributing to the president the belief that Stanley P. Young is a spy and thus a sufficient basis for taking this to be true:

(7) The president believes that Stanley P. Young is a spy.

If we construe 'The president believes…is a spy' as a complex predicate, then from (7) and the law of Existential Generalization, we should be able to infer:

(8) $(\exists x)(\text{The president believes that } x \text{ is a spy})$.

There may be a number people whom the president believes to be spies. Stanley P. Young is one such person. $(8')$ below would also seem to follow:

$(8')$ $(\exists x)(x = \text{Stanley P. Young \& the president believe that } x \text{ is a spy})$.

Now suppose that the president is subsequently introduced to one James Q. Money, a generous contributor to progressive causes and a man of stellar reputation. The president trusts Mr Money. When asked whether Money is a spy, the president denies even the possibility. That denial is grounds for inferring:

(9) The president believes that James Q. Money is not a spy.

From (9) and Existential Generalization we can infer:

(10) $(\exists x)(\text{The president believes that } x \text{ is not a spy})$.
$(10')$ $(\exists x)(x = \text{James Q. Money \& the president believe that } x \text{ is not a spy})$.

But unbeknownst to the president, James Q. Money leads a double life. He is none other than Stanley P. Young. Does the president believe or not believe James Q. Money, that is, Stanley P. Young, to be a spy? Intuition pulls in opposite directions. Since James Q. Money just is Stanley P. Young, we should, by Leibniz's law, be able to substitute 'Young' for 'Money' in $(10')$ without change of truth-value. Since this substitution happens outside the that-clause, it should be unaffected by the logical peculiarities of such clauses. Yet substitution yields:

$(10'')$ $(\exists x)(x = \text{Stanley P. Young \& the president believe that } x \text{ is not a spy})$.

Examples $(10'')$ and $(8')$ are of dubious consistency. No one can both be and not be a spy. So how can anyone consistently both believe a person to be a spy and believe that very person not to be a spy?

The answer lies in referential opacity. Because (7) and (9) are opaque, they do not ascribe to the president incompatible relations to Money, that is, Young. Indeed, opaquely construed, the Quinean insists, neither (7) nor (9) relates the president to Money, that is, Young at all. Recall again the Quinean view that when opaque contexts are perspicuously spelled the illusion that they contain referring expressions as proper constituents will have been dispelled. Indeed, opacity renders quantified sentences like $(10')$ and $(10'')$ utterly nonsensical. For proper respelling reveals that there is no genuine occurrence of an interior variable for the exterior quantifier to bind. An analogy with quotation - which Quine takes to be the referentially opaque context *par excellence* - makes the reason clear. Consider:

(11) 'x is a spy' is an open sentence.

Enclosing an expression within quotation marks seals that expression off from the reach of external quantifiers. The quantifier below does not bind the quoted variable:

(12) $(\exists x)$('x is a spy' is an open sentence).

What occurs within the quotation marks is really not a variable *in use*. It is rather a variable *being mentioned* (see

Use/mention distinction and quotation). Similarly, though what occurs in $(10')$ and $(10'')$ bears a certain orthographic resemblance to a variable in use, it is not the real thing. For that reason we can no more quantify into propositional attitude contexts, Quine insists, than we can quantify into the context of quotation.

Quine's approach promises a way out of the threatened contradiction, but only by doing great violence to a number of potent intuitions. It seems frankly incredible that putatively referring terms embedded within a that-clause are mere grammatical illusions. Further, if propositional attitude statements lack logically germane constituents, there will be no accounting for certain systematic commonalities and differences among such statements. Compare and contrast, for example, (7) and (9) above with (13) and (14) below:

(13) The First Lady *doubts that* Young is a spy
(14) The vice-president *wonders whether* Young is a spy.

Example (13) seems to say that the First Lady doubts what the president believes, while (14) says that the vice-president has questions about what the First Lady and the president disagree about. Moreover, the president's belief, the First Lady's assurance and the vice-president's questions seem all to be about one and the same individual. Quine's approach abandons these intuitive judgments.

Quine is well aware of these costs. He believes, however, that we must pay the cost if we are to keep the creatures of darkness - Fregean senses and other intensional entities - at bay (see Intensional entities).

3 Innocence regained?

Donald Davidson has argued that if we could but regain our pre-Fregean semantic innocence:

> it would seem to us plainly incredible that the words 'The earth moves' uttered after the words 'Galileo said that' mean anything different, or refer to anything else, than is their wont when they come in other environments.

<div align="right">(Davidson 1969: 172)</div>

If Davidson is right then we can account for the logical peculiarities of attitude statements without positing Fregean reference shifts. Nor, if he is right, need we allow that otherwise referring terms are stripped of their referential functions when they occur in attitude contexts. Indeed, against Quine, Davidson holds that unless we have already been forced to introduce Fregean senses for other reasons, the peculiarities of attitude contexts provide no additional pressures to posit such entities. So if Davidson is right, Quine's drastic departures are not needed to keep intensions at bay.

Davidson defends what he calls a 'paratactic analysis' of the logical forms of indirect discourse, though his approach can be extended fairly directly to attitude ascriptions. Davidson's central claim is that a sentence like:

(15) Galileo said that the earth moves

is really a parataxis of two sentences. The first sentence of the parataxis in (15) ends with the word 'that'; the second starts with 'the earth'. By adding a bit of punctuation, we can represent (15) in a more perspicuous manner as:

(16) Galileo said that. The earth moves.

Call the first sentence the attribution sentence and the second the content sentence. In the attribution sentence of (16), the 'that' is a demonstrative pronoun which refers (on an occasion) to an utterance of the second sentence of (16).

Sentence (16) will be true, according to Davidson, just in case some utterance of Galileo's makes Galileo and the utterer of (16) 'samesayers'. Suppose that Galileo utters (17) and Smith utters (18):

(17) Eppur si muove.
(18) The earth moves.

In some sense, Galileo and Smith say the same thing. Moreover, if Smith utters (18) and Galileo utters (17) then

'the earth' in Smith's mouth and 'Eppur' in Galileo's mouth have the same reference.

Suppose that Smith wants to attribute Galileo's statement to Galileo. He can do so simply by saying the same thing as Galileo and then saying that he has just done so - as in the following scenario:

(19) Galileo: Eppur si muove.
(20) Smith: The earth moves. Galileo said that (too).

Here we have Galileo saying, in Italian, that the earth moves. We have Smith saying the same thing in English. And then we have Smith ascribing such a statement to Galileo. He does so by making a statement that turns himself and Galileo into samesayers (to use Davidson's phrase) and then stating that he has just done so. As matters stand, it remains unclear until Smith utters the second sentence of (20) just what the point of his initial utterance may be. But we can rectify that by reversing the order of the sentences as follows:

(21) Galileo said that. → The earth moves.

The arrow in (21) should be understood as a demonstration accompanying the demonstrative pronoun 'that'. The demonstrative refers to Smith's second utterance - the one by which Smith purports to turn himself and Galileo into samesayers. While Smith 'refers' to his own utterance, he implicitly 'quantifies over' Galileo's utterances. Smith says, in effect, that the demonstrated utterance of his and 'some utterance or other' of Galileo's make himself and Galileo samesayers.

Davidson's account yields a relatively straightforward explanation of the invalidity of inferences like the following:

(22) Galileo said that. → The earth moves.
(23) The earth is the third planet from the sun.
(24) Galileo said that. → The third planet from the sun moves.

There is no logical connection between the effect of substitution on the content sentence and the truth-values of the two attribution sentences. As Davidson puts it:

> There is no reason to predict, on grounds of form alone, any *particular* effect on the truth of [the attribution sentence] from a change in [the content sentence]. On the other hand, if the [content sentence] had been different in any way at all, [the attribution sentence] *might* have had a different truth-value, for the reference of the 'that' would have been changed.
>
> (Davidson 1969: 172)

Though Davidson's central idea is quite ingenious, there are both syntactic and semantic grounds for scepticism. For example, content clauses are not generally introduced by demonstratives. Neither 'whether' in 'Galileo wondered whether the earth moves' nor 'for… to' in 'I would prefer for you to leave' can plausibly be analysed as overt demonstratives. Moreover, demonstratives cannot in general be deleted, but the 'that' in '… said that… ' can. Compare the following:

(25) Davidson saw that man over there.
(26) Davidson said that the paratactic analysis is true.
(27) ?Davidson saw man over there.
(28) Davidson said the paratactic analysis is true.

It is also difficult to see how to extend the paratactic analysis to *de re* ascriptions. Consider (29):

(29) Every philosopher believes that he is wise.

Example (29) says, in effect, that every philosopher has a high opinion of himself. But if we interpret (29) along Davidsonian lines as:

(30) Every philosopher believes that. He is wise.

we get something rather different.

Finally it is clear that we need to know more about the samesaying relation. If the paratactic analysis is to overcome the need to introduce Fregean senses and the like, then samesaying cannot just be a matter of expressing the same Fregean sense. Davidson owes us an account of samesaying which eschews senses and other intensional entities altogether. Davidson is fully aware of this debt and labours mightily to pay it (Davidson 1969).

4 Structured complexes

A class of theories which has recently gained favour with a number of philosophers is what might be called the structured complex approach (see Intensional logic §5). Advocates of such theories maintain that attitude verbs express relations between agents and complex hierarchically structured entities of one sort or another. The core idea dates back to Carnap (1947). It has resurfaced, much modified, in Cresswell (1984), Segal (1989), Richard (1990), Higginbotham (1991) and Larson and Ludlow (1993). Different versions of this approach posit different structured complexes. Carnap posited hierarchically structured intensions. Others pair semantic values with something else: sometimes a sentence, sometimes a description of a sentence, sometimes a structured intension. Larson and Ludlow, for example, hold that that-clauses specify Interpreted Logical Forms (ILFs). An ILF is a kind of product of a phrase structure tree for a sentence and the sequence of semantic values for the semantically valued constituents of that sentence. On their approach, the sentence:

(31) Pierre admires London

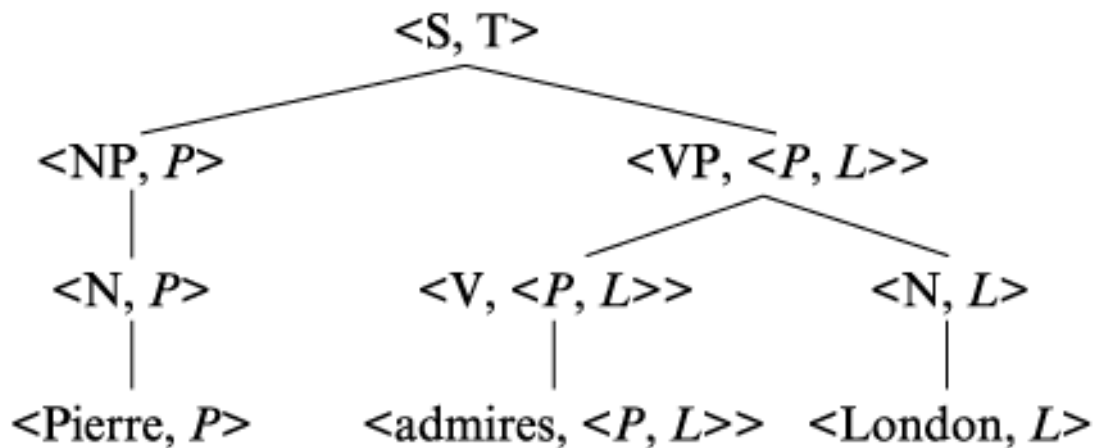will be associated with the ILF in Figure 1:

Figure 1

We need not worry about the exact formation rules that generate this ILF. The crucial point is that each node in this tree is labelled by a pair whose left member is either a lexical item (at the bottom-most nodes) or a phrase marker (at all other nodes) and whose right member is a semantic value. For example, the pair $\langle \text{Pierre}, P \rangle$ is the pair whose left member is the word 'Pierre' and whose right member is the reference of 'Pierre' (namely, Pierre himself). The pair $\langle \text{admires}, \langle P, L \rangle \rangle$ has left member 'admires' and a right member consisting of a pair whose members are Pierre and London. Such trees are intended to represent a sequence of semantic values and the order in which those semantic values are associated with the expressions whose values they are. Because lexical items (words themselves), and not just phrase makers, are constituents of ILFs, the following two sentences will have distinct ILFs:

(32) Hesperus rises in the evening.
(33) Phosphorus rises in the evening.

Nonetheless, (32) and (33) have, according to Larson and Ludlow, the same semantic content. That is, (32) and (33) have formally distinct but semantically equivalent ILFs. Armed with the notion of formally distinct but semantically equivalent ILFs, Larson and Ludlow are able to explain how it is possible for attitude statements like our original (2) and (3) to differ in truth-value. The thought is that an agent who stands in the believing relation to

the ILF of 'Hesperus rises in the evening' need not *ipso facto* stand in the believing relation to the ILF of 'Phosphorus rises in the evening.' Yet one who believes that Hesperus rises believes the very same semantic content as one who believes that Phosphorus rises.

Notice that Larson and Ludlow walk the middle ground between Davidson and Frege. Their approach is ontologically conservative like Davidson's. For the structural descriptions and semantic values out of which ILFs are composed must already be introduced in to our ontology to account for the semantic and syntactic character of sentences other than those about propositional attitudes. But their explanation of failures of substitutivity is analogous to Frege's own. They too appeal to the fact that, at least in attitude contexts, more is relevant to substitutivity than mere co-reference. That more is, however, not anything semantic. So we are not required, merely because of the peculiarities propositional attitude statements, to posit for each referring expression a second semantic role over and above its role of standing for a referent.

5 Hidden indexicality

Consider the following three theses:

(a) *The Direct Reference Thesis.* The sole semantic role of a proper name is to stand for its bearer.
(b) *Semantic Innocence.* A name plays the same semantic role when embedded within a that-clause as it plays when it occurs outside a that-clause.
(c) *The Fregean Intuition.* Propositional attitude statements which differ only by co-referring proper names may differ in truth-value.

Each of these theses has considerable independent plausibility. But they are not obviously consistent. Frege was moved to abandon both thesis (a) and (b) because he thought they could not be reconciled with (c) (see Proper names §5). More recently, Nathan Salmon (1986) has abandoned (c) out of a commitment to (a) and (b). He has insisted that it is merely a pragmatically generated illusion that sentences like (2) and (3) have different truth-values. He claims that strictly, literally speaking, one who believes that Hesperus rises *ipso facto* believes that Phosphorus rises. It is just that it is conversationally misleading to talk that way. That is because belief is not a two place relation between a believer and a proposition, but a three place relation between a believer, a proposition and what we might call a 'way of believing' that proposition. Propositions are never believed *simpliciter*, but always in some way or other, under some guise or other. Moreover, Salmon insists, a rational cognizer can believe a given proposition under one guise while failing to believe it under a different guise. But the crucial further point is that nothing in an attitude ascription semantically specifies the guise under which the proposition named by the that-clause is supposed to be believed. The guise simply does not enter into the strict literal truth-conditions of an attitude statement. Nonetheless, Salmon suggests, we can pragmatically suggest something about the guise, by using one expression rather than another semantically equivalent one in making an attribution. If we use 'Hesperus' we suggest something different about the guise than we would if were to use 'Phosphorus'. That is why it is pragmatically misleading to use 'Hesperus' and 'Phosphorus' interchangeably in making attitude ascriptions.

By contrast, hidden indexicalists insist on all of our principles, and offer a strategy, with several variants, for reconciling them (see Schiffer 1979; Crimmins and Perry 1989; Richard 1990; Crimmins 1992). Proponents of the hidden indexical approach generally agree with the view that belief is a three-place, rather than merely a two-place relation. They even agree that that-clauses which differ only by co-referring proper names must refer to the same (singular) proposition. But they insist that attitude ascriptions do not just pragmatically suggest information about the guise under which the specified proposition is believed. Information about the guise is part of what is strictly, literally said by an attitude ascription. That is why, against Salmon, the Fregean intuition is here to stay.

Yet, given the assumption of semantic innocence, how could that-clauses which differ only by corresponding co-referring proper names specify distinct ways of believing? The answer, according to the hidden indexicalist, is that there is more to an attitude statement than meets either the eye or the ear. Attitude reports contain hidden or unarticulated indexical constituents. These unarticulated constituents are not reflected in the surface grammatical structure of the sentence, but in a context of utterance they somehow refer to a way of believing. Consequently, attitude sentences, in context, manage to express propositions with more explicit constituents than they themselves have. Such sentences can be compared, the hidden indexicalist says, to sentences like the following in which one term of a relation is suppressed, but can be supplied, in context, by competent producers and consumers:

It is raining.   (Where?)
I am ready.   (For what?)
That is tall.   (For a what?)
Smith believes that *p*.   (How?)

When one utters, 'It is raining' on a certain occasion, one strictly, literally says that it is raining in a certain place - typically the place of utterance - even though no explicit constituent of one's utterance refers to that place. Just so, the hidden indexicalist maintains, when one utters a belief report on an occasion, one strictly, literally says that a certain proposition is believed in a certain way. And one does so even though no explicit constituent of one's utterances refers to the way of believing. Now the crucial further claim has to be that although 'Hesperus' and 'Phosphorus' have the same semantic value and function innocently, even in that-clauses, nonetheless, exchanging one for the other can affect the way an unarticulated constituent gets hooked up to a reference as a function of context. And that can explain why they are not freely substitutable one for another.

Clearly, the hidden indexicalist owes us a detailed account of these matters. We need to be told what a way of believing is. We need to know how specified ways of believing vary as a function of facts about context and facts about articulated constituents. Fortunately, the best advocates of this approach have gone a long way towards answering such questions.

See also: Demonstratives and indexicals; Indirect discourse; Propositional attitudes

KENNETH A. TAYLOR

References and further reading

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press; enlarged 2nd edn, 1956. (Introduces the notion of an intensional isomorphism, the earliest version of the structured complex approach.)

Cresswell, M. (1984) *Structured Meanings*, Cambridge, MA: MIT Press.(Argues for structured intensions as objects of propositional attitudes.)

Crimmins, M. (1992) *Talk about Beliefs*, Cambridge, MA: MIT Press.(Defends a hidden indexical theory.)

Crimmins, M. and Perry, J. (1989) 'The Prince and the Phone Booth: Reporting Puzzling Beliefs', *Journal of Philosophy* 60: 435-45.(Defends a hidden indexical theory.)

Davidson, D. (1969) 'On Saying That', in D. Davidson and J. Hintikka (eds) *Words and Objections: Essays on the Work of W.V. Quine*, Dordrecht: Reidel, 158-74. (Articulates the paratactic analysis of indirect discourse.)

Davidson, D. (1967) 'Truth and Meaning', *Synthèse* 17: 304-23.(Articulates Davidson's broader semantic programme.)

Devitt, M. (1996) *Coming to our Senses*, Cambridge: Cambridge University Press. (Defends a neo-Fregean account of attitude statements without positing Fregean senses; rejects hidden indexical approaches and direct reference.)

Fodor, J. (1989) 'Substitution Arguments and the Individuation of Beliefs', in G. Boolos (ed.) *Method, Reasons, and Language*, Cambridge: Cambridge University Press.(Construes modes of presentations as expressions in the language of thought, while holding a referentialist semantics.)

Frege, G. (1892) 'Über Sinn und Bedeutung' (On Sense and Reference), *Zeitschrift für Philosophie und philisophische Kritik* 100: 25-50.(A seminal work of Frege's in which he argues for the distinction between the sense and reference of expressions and presents a number of important theses regarding the two notions. Translated as 'On Sense and Meaning' in P.T. Geach and M. Black (eds) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 3rd edn, 1980.)

Higginbotham, J. (1991) 'Belief and Logical Form', *Mind and Language* 6: 344-69. (Defends a version of the structured complex approach, drawing on the resources of contemporary syntactic theory. Relates this approach to Davidsonian extensional semantics.)

Larson, R. and Ludlow, P. (1993) 'Interpreted Logical Forms', *Synthèse* 95: 305-57.(Articulates and defends the ILF approach.)

Quine, W.V. (1956) 'Quantifiers and the Propositional Attitudes', repr. in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press, revised and enlarged edn, 1976.(This widely reprinted article contains Quine's argument against the permissibility of quantifying into opaque contexts.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: Harvard University Press. (Forcefully articulates Quine's disdain for intensions as 'creatures of darkness'.)

Recanati, F. (1993) *Direct Reference*, Oxford: Blackwell.(A full-scale defence of the direct reference thesis which articulates yet another way to reconcile direct reference, innocence and the Fregean intuition.)

Richard, M. (1990) *Propositional Attitudes: an Essay on Thoughts and How We Ascribe Them*, Cambridge: Cambridge University Press.(Argues that propositional attitude statements express relations to Russellian annotated matrices - a version of the structured complex approach; also defends the view that belief reports are indexical.)

Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.(Denies the Fregean intuition and defends innocence and direct reference.)

Schiffer, S. (1979) 'Naming and Knowing', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press. (First appearance of the hidden indexical approach.)

Segal, G. (1989) 'A Preference for Sense and Reference', *Journal of Philosophy* 86: 73-89.(Another defence of a structured complex approach.)

Soames, S. (1987) 'Direct Reference, Propositional Attitudes, and Semantic Content', *Philosophical Topics* 15: 47-87.(Denies the Fregean intuition, while defending innocence and direct reference.)

Propositional attitudes

Examples of propositional attitudes include the belief that snow is white, the hope that Mt Rosea is twelve miles high, the desire that there should be snow at Christmas, the intention to go to the snow tomorrow, and the fear that one shall be killed in an avalanche. As these examples show, we can distinguish the kind of attitude - belief, desire, intention, fear and so on - from the content of the attitude - that snow is white, that there will be snow at Christmas, to go to the snow, and so forth. The term 'propositional attitudes' comes from Bertrand Russell and derives from the fact that we can think of the content of an attitude as the proposition the attitude is towards. It can be typically captured by a sentence prefixed by 'that', though sometimes at the cost of a certain linguistic awkwardness: it is more natural, for example, to talk of the intention to go to the snow rather than the intention that one go to the snow. The most frequently discussed kinds of propositional attitudes are belief, desire and intention, but there are countless others: hopes, fears, wishes, regrets, and so on.

Some sentences which contain the verbs of propositional attitude - believes, desires, intends, and so on - do not make ascriptions of propositional attitudes. For example: 'Wendy believes me', 'John fears this dog', and 'He intends no harm'. However, while these sentences are not, as they stand, ascriptions of propositional attitudes, it is arguable - though not all philosophers agree - that they can always be analysed as propositional attitude ascriptions. So, for example, Wendy believes me just in case there is some p such that Wendy believes that p because I tell her that p; John fears this dog just in case there is some X such that John fears that this dog will do X and so on.

Discussions of propositional attitudes typically focus on belief and desire, and, sometimes, intention, because of the central roles these attitudes play in the explanation of rational behaviour. For example: Mary's visit to the supermarket is explained by her desire to purchase some groceries, and her belief that she can purchase groceries at the supermarket; Bill's flicking the switch is explained by his desire to illuminate the room, and his belief that he can illuminate the room by flicking the switch; and so on. It is plausible - though not uncontroversial - to hold that rational behaviour can always be explained as the outcome of a suitable belief together with a suitable desire.

Some philosophers (examples are Grice and Schiffer) have used the propositional attitudes to explain facts about meaning. They hold that the meanings of sentences somehow derive from the contents of relevantly related beliefs and intentions. Roughly, what I mean by a sentence S is captured by the content of, say, the belief that I express by saying S.

One fundamental question which divides philosophers turns on the ontological status of the propositional attitudes and of their contents. It is clear that we make heavy use of propositional attitude ascriptions in explaining and interpreting the actions of ourselves and others. But should we think that in producing such ascriptions, we attempt to speak the truth - that is should we think that propositional attitude ascriptions are truth-apt - or should we see some other purpose, such as dramatic projection, in this usage? Or, even more radically, should we think that there is nothing but error and confusion - exposed by modern science and neurophysiology - in propositional attitude talk?

1 Elementary distinctions

Propositional attitude ascriptions standardly take the form 'X Fs that p', where 'X' denotes the subject of the attitude, 'F' is a verb of propositional attitude, and 'that p' gives the content of the attitude. So, for example, in the sentence 'John believes that snow is white', 'John' denotes the subject of the attitude, 'believes' is the verb of propositional attitude, and 'that snow is white' gives the content of the attitude.

Taking the form of these ascriptions at face value, it is natural to suggest that a propositional attitude ascription of the form 'X Fs that p' is true just in case X is in a propositional attitude state of type F - a F-state - which has the content that p. So, for example, 'John believes that snow is white' is true just in case John is in a belief-state which has the content that snow is white; 'Mary hopes that Sheila is happy' is true just in case Mary is in a hope-state which has the content that Sheila is happy; and so on. More generally, it is natural to divide a discussion of propositional attitudes into two parts: one part which focuses on the nature of propositional attitude states - the difference between hoping, fearing, believing, wishing, and so on - and one part which focuses on propositional

attitude contents - the difference between believing that snow is white, believing that snow is grey, believing that snow falls at night, and so forth.

It is plausible to think that belief and desire are the basic exemplars of two quite different kinds of attitudes, with characteristically different directions of fit. On the one hand, there are attitudes, like belief, which aim to fit the world - and, hence, which are important for theories of truth, impact of evidence, credence, and so on. And, on the other hand, there are attitudes, like desire, which aim to have the world fit them - and, hence, which are important for theories of value, virtue, wellbeing, and so on. Some philosophers think that all of the propositional attitudes can be explained in terms of beliefs and desires; they think that the apparent multiplicity of kinds of propositional attitudes is merely apparent - hope, for example, is a kind of desire about the future - and that direction of fit is the only fundamental dimension which needs to be considered in classifying the propositional attitudes.

2 Propositional attitude states

If beliefs, desires, intentions, and the like, are states of subjects, what kind of states are they? A plausible view is that they are functional states - states defined or determined or specified by their functional roles. According to functionalist theories, propositional attitude states are fully determined by their causal relations to one another, to perceptual inputs, and to behavioural outputs (see Functionalism). So, for example, what makes a certain state a state of believing that there are *Cheerios* in the cupboard turns on its being typically caused by packets of *Cheerios*, or more precisely, on the role it plays in the processing of perceptual input - say, guessing the contents of that visually presented cardboard box in the cupboard; on the way it leads to other states - say, believing that there is cereal in the cupboard; and on the way it combines with hunger and desire to eat some cereal to produce behaviour that leads to eating.

There are, however, two different ways of thinking of functional states: as role states or as realizer states. Realizer states are first-order states that stand in the functional roles; they are the particular states which play the roles. Role states are second-order states: the state of being in, some first-order state playing a certain kind of functional role. So, in the case of belief, say, there is the first-order realizer state (the state which actually fills the belief role - the state which mediates between perceptual inputs and behavioural responses in the way distinctive of belief); and there is the second-order role state (the state of being in some state or other which realizes the belief role - the state of being in a state that mediates in the way distinctive of belief between perceptual inputs and behavioural outputs). Should we think of belief in particular and propositional attitudes in general as realizer states or as role states?

The obvious reason for holding that propositional attitude states are realizer states rather than role states is that the propositional attitudes are centrally involved in the causation of behaviour, and it is the realizer states but not the role states that are centrally involved in the causation of behaviour. It is, for instance, the particular internal state which fills the role we associate with believing that snow is white that causes the behaviour distinctive of this belief - for example, uttering the words 'Snow is white'.

There are, however, also considerations that point the other way, that favour holding, for example, that my belief that snow is white is the role state which is realized by some particular internal state of mine. In particular, it seems that taking propositional attitude states to be role states allows one to formulate psychological generalizations which one would miss if one took propositional attitudes to be realizer states. Consider, for example, the generalization: everyone who believes the world is going to end in ten minutes gets anxious. Suppose that there are actually M realizer states for the belief: $N_1, N_2, \ldots, N_M$. These states, we may suppose, each play the appropriate causally intermediate roles distinctive of the belief that the world is going to end in ten minutes in somewhat the way that transistors and valves, despite being very different, can both play the roles distinctive of amplification. If the belief state is identical with the realizer states, then the only way to generalize is to treat the realizer state as a disjunction $N_1 \vee N_2 \vee \ldots \vee N_M$. But this disjunction is ugly, and not suitable for the formulation of psychological law. The generalization can be saved by recasting it as one about role states, but then the only way the initial generalization can be a generalization about the belief that the world is going to end in ten minutes is if that belief is a role state. The general point is that the patterns we capture in psychological generalizations relate to role states, so the cost of holding that psychological states are realizer states is that psychological generalizations cease to relate psychological states.

Whether propositional attitude states are best thought of as role states or realizer states, it has seemed plausible to many philosophers that the realizer states are physical states of the brain - or, at any rate, states which supervene upon physical states of the brain (see Mind, identity theory of; Supervenience of the mental). This is put forward as a plausible scientific hypothesis. Only in the brain are there states of sufficient complexity to play the needed roles. It is, though, allowed as a logical possibility that there could have been other (perhaps even non-physical) realizer states which filled the same role.

Some philosophers, however, have worried about whether neural states could be the kinds of things which can be adverted to, even *inter alia*, in the explanation of rational action. The connections between neural states are merely causal, they argue - playing out the universal laws of physics, or of neuroscience - whereas the connections between states of propositional attitude are rational (and serve to rationalize behaviour). One belief makes another the rational one to have; what one believes and desires may make one action more rational than another; and so on (see Reasons and causes).

However, there are serious problems with this sort of objection. Would the cause of rationality be better served by having the attitudes occurring at random? And surely a decent account of the propositional attitudes places them in the physical world - they are not otherworldly mysteries - and given this, there seems little alternative to placing them in the brain: we know that the propositional attitudes play complex causal roles in the production of our behaviour, and only the brain has states capable of the needed complexity. Finally, it is unclear why one might suppose that functionally defined states cannot have the rationalizing properties which are alleged to vanish from the scientific picture (see Loar 1981 for further discussion).

3 Propositional attitude contents

The contents of propositional attitudes are propositions. That much is simply a matter of definition. But what are propositions? There are at least the following contenders: (1) collections of circumstances of evaluation, such as collections of possible worlds; (2) syntactic entities, such as interpreted sentences in natural language; (3) set-theoretic structures, such as Fregean thoughts (see Frege §§3-4). Only some of the strengths and weaknesses of these different approaches can be examined here (see Propositions, sentences and statements).

(1) One important virtue of the theory that propositions are collections of possible worlds - that is, collections of complete ways that things could be (see Semantics, possible worlds) - is that it directly captures the intuition that propositions involve a sorting among ways that things could be. For example, in believing that snow is white, one believes that the actual complete way that things are is among those complete ways that things could be in which snow is white; in desiring that snow falls at Christmas, one desires that the actual complete way that things are is among those in which snow falls at Christmas; and so on.

Perhaps the most important drawback of this theory is that its most plausible formulations seem to give the wrong results for attitudes with logically equivalent contents. For example, it entails that the belief that there are husbands is identical to the belief that there are wives. Because 'There are wives' is logically equivalent to 'There are husbands', the set of worlds where there are wives is one and the same as the set of worlds where there are husbands. Likewise, it makes the desire to prove that two and two are four identical to the desire to prove that arithmetic is not decidable, and the intention to draw an equiangular triangle identical to the intention to draw an equilateral triangle. As the objection is commonly put, the possible worlds theory delivers objects of the attitudes that are insufficiently fine grained - there are more distinct attitude contents than it allows.

In addition, there is a complication which needs to be addressed. Consider, for example, my belief that I am cold, or your belief that you are spilling sugar on the floor of the supermarket. These attitudes *de se* (about oneself), as they are called, need to be thought of as involving a sorting among ways that things might be centred on the subject - me, here and now, as it could be, rather than of complete ways things might be (see Demonstratives and indexicals; Content, indexical).

(2) One important virtue of the theory that propositions are syntactic entities - that is entities with sentence-like structure - is that it provides objects to discriminate between attitude contents which, as we have just seen, the possible worlds theory is unable to discriminate. For example, the sentence 'There are husbands' is distinct from the sentence 'There are wives'; and the theory can exploit this fact to distinguish the belief that there are husbands from the belief that there are wives. They are attitudes to different though equivalent sentences.

However, if the sentences are thought of as being in a natural language, this advantage is purchased at a price; it provides distinctions where it is plausible that there are none. Consider, for example, our practice of translating and interpreting those who speak different languages. Clearly, 'Snow is white' and 'La neige est blanche' are distinct sentences - but surely we are not thereby obliged to say that it is impossible for a monolingual French speaker to believe that snow is white? Consider, too, the belief that p, and the belief that p and p. There is at least some temptation to think that these are not distinct beliefs.

One response is to argue that the sentences in question do not belong to natural languages but rather to a language of thought - that is, to a neural system of information or information storage which is supposed to have syntactic structure, and that this is something we all share (and share with animals like dogs that lack public language but do seem to have propositional attitudes) (see Language of thought). One difficulty with this suggestion is that it seems wrong that a mere analysis of propositional attitudes will commit us to the existence of a language of thought, a substantive view about our neural natures. It seems to be an empirical question, for example, whether mental representation is language-like or map-like (see Belief §3), but the language of thought hypothesis commits us in advance to the former alternative.

(3) One important virtue of the theory that propositions are set-theoretic structures - that is entities whose basic structure is described by set theory (see Set theory) - is that it promises to cope with the difficulties mentioned above for the syntactic and possible worlds accounts. However, before we can see why this is so, we need to give some explanation of the theory.

Consider, for example, the sentence 'London is pretty'. A natural thought is that the proposition expressed by this sentence is somehow made up of the city of London (or some way of thinking about it) and the property of being pretty (or some way of thinking about the property of being pretty). More generally, the theory holds that propositions are composed from constituents which are somehow put together in an orderly (set-theoretic) fashion. The constituents in question include: individual objects (and perhaps modes of presentation thereof), which correspond to proper names and other singular terms; properties, which correspond to predicates; functions, which correspond to sentential operators (such as 'and' and 'possibly'), and so on.

This view has the resources to discriminate between the belief that there are husbands and the belief that there are wives, one of the examples that gave the possible worlds approach trouble. Roughly, the first has in its propositional object women and the property of being married to men, whereas the second has men and the property of being married to women. The view also has the resources to identify the propositions expressed by the sentences 'Snow is white' and 'La neige est blanche' and so allow that speakers of different languages can share beliefs, one of the examples that gave the sentential theory trouble. Although the words differ, the objects and properties they stand for are the same and hence the propositions thought of as constructed from them are the same. However, the theory in its typical form has to distinguish between the belief that p, and the belief that p and p - the propositional object of the second is special in involving the operation of conjunction - and this is not clearly desirable.

Another problem with set-theoretic approaches is that they involve various kinds of 'disreputable entities' - intensional properties, modes of presentation ('Fregean senses'), perhaps even set-theoretic functions - which many philosophers would like to exclude from their theories of the attitudes (see Intensional entities). They worry about how these entities can be tied to propositional attitudes, especially if these attitudes are thought of functionally. What causal role can intensional properties or sets, for instance, play? True, this is part of a hard question that arises for any account of the attitudes. We have been discussing competing views about what the objects of the attitudes are. But there is also the question as to how a given propositional attitude state gets attached to the proposition it is attached to; the question of what makes it true, for example, that a certain state is the belief that snow is white rather than that grass is green. We will set this important question aside except to remark that the set theoretic view of the nature of the attitudes is often thought to be especially ill placed to tackle it.

Plainly, there is much more to be said about each of the options discussed here (and about other options as well). The important point is that there are many dimensions to propositional content, and that it is far from easy to provide a theory which scores well on every dimension. A good theory of propositional content should have all of

the following features:

(1) it should be naturalistically respectable - that is compatible with the outlook of natural science;
(2) it should capture the normative dimensions of propositional content;
(3) it should allow a fine-grained discrimination of the contents of propositional attitudes (logical equivalence is not sufficient for identity);
(4) it should not tie the possession of propositional attitudes too tightly to particular languages (for example, it should allow that the same attitudes can be had by creatures which do not share a language - and, indeed, that attitudes can be possessed by creatures which do not speak any language at all);
(5) it should make it clear how states which possess that kind of propositional content can play a role in the production of (rational) action.

4 Propositional attitude 'aboutness'

Many philosophers have been puzzled by the various ways and senses in which propositional attitudes can be about particular objects and kinds of objects (see Intentionality). Think, for example, of the beliefs which one might have about Aristotle, about Einstein, about President Clinton, about Zeus, about Mickey Mouse, about Meinong's round square, about gold, about phlogiston, about π, and about the greatest prime number. There are several different issues which can arise here.

One issue concerns the existence of the objects and kinds of objects which are the focus of the attitudes. There is a sense of 'about' in which one cannot have attitudes about an object (or kind of object) which does not exist - so that, for example, in this sense, one cannot have beliefs, hopes, fears, desires, and so on, about Zeus, the round square, the greatest prime number, Mickey Mouse and phlogiston, though one can have these attitudes about President Clinton and gold. (The cases of entirely past objects - Aristotle and Einstein - and numbers - π- raise other controversial questions about existence which cannot be taken up here.) This sense of 'about' is relational: the point parallels the impossibility of parking near the round square, shaking hands with Zeus, or adding three to the greatest prime number.

Another issue concerns the nature of the connection which obtains between subjects and those existent objects and kinds of objects which are the focus of given attitudes. If one is not appropriately connected to an object, then - in this sense - one cannot have attitudes about it. Suppose, for example, that by an extraordinary coincidence there is a planet somewhere in the universe whose history is exactly described by Tolkien's *Lord of the Rings*. Even though there is some sense in which Tolkien's work is true of this planet, nonetheless, the work is not about the planet (in the sense currently at issue). The reason is that the planet played no role at all in the causal history of the book. When, for example, one believes that Hawking is a genius, what makes it the case that one's belief is about Hawking is, very likely, at least in part, that there are chains of usage of the word 'Hawking' which are appropriately connected to that theoretical physicist. When I believe that that tree outside my window is shedding its leaves, what makes it a belief about that tree is, at least in part, the causal processes involving electromagnetic radiation passing between the tree and my retina which are involved in the formation of my belief. In general, it seems plausible to think that some kind of causal connection is required for these kinds of attitudes about particular objects.

However, while it is clearly correct to say that there is a sense in which the aboutness of propositional attitudes is a matter of standing in a relation, an appropriate causal connection on most views, it is also clear that there is another sense of 'about' which cannot be analysed in this way. In some sense we can have beliefs and desires about non-existent objects - for example, some people desire to find the fountain of youth, some believe that Meinong's round square is round, some hoped that Zeus would smite their enemies with a thunderbolt, and so on. It seems clear that those who, for example, hoped that Zeus would smite their enemies have a hope 'about' Zeus despite not standing in a relation, causal or otherwise, to Zeus.

How then should this final, non-relational, non-causal, sense of 'about' be analysed? Many alternatives present themselves. One might appeal to counterfactuals: attitudes 'about Zeus' are attitudes about something which does not exist, but which might have existed; attitudes 'about phlogiston' are attitudes about something which does not exist but which might have existed; and so on. One might hold that this sense of 'about' is only used in loose speech, and can be banished or paraphrased away in serious philosophical discourse: attitudes 'about Zeus' are

really attitudes which focus on certain myths and stories; attitudes 'about phlogiston' are really attitudes about a false chemical theory; and so on. One might try various other options as well. The issue is too difficult to discuss adequately here: the important point is that there is at least a loose sense in which there is 'existentially neutral' content, that is, content which does not commit one to the existence in the actual world of the particular individuals and kinds to which the content does give apparent commitment.

There is also an important distinction between two different kinds of content: broad content and narrow content. Roughly speaking, narrow content is purely a matter of what goes on inside the head of a subject of propositional attitude states, whereas broad content is in part a matter of how things are in the world outside the head. More exactly, if two subjects are intrinsic duplicates who live under the same laws of nature and belong to the same kinds, then their attitudes are exactly alike in narrow content but may differ in broad content (see Content: wide and narrow).

Once upon a time it was more or less taken for granted that the content of belief, for example, was narrow. How things are outside the head obviously affects whether your beliefs are true or false - your belief that London is pretty can only be true if London itself is pretty - but what you believe is a matter of how you are. But in fact the belief that London is pretty is a belief with broad content. I cannot believe that London, the capital city of England, is pretty unless I am appropriately causally connected to London, and that concerns how things are outside my head. Imagine that the universe is symmetrical, so that there is a region - call it Twin Earth - which is qualitatively identical to our own at all times. It will contain a city called 'London' and a twin of me. This twin will be an intrinsic duplicate of me and will produce the sentence 'London is pretty' as expressing what he believes. Nevertheless, his belief (and his sentence) will have a different content from mine. It will be about Twin London rather than about London, and will be true just if Twin London, rather than, London is pretty.

5 Defining propositional attitudes

The characterization of propositional attitudes in §2 was very rough, and leaves room for the thought that functionalism involves circularity of definition. We in effect identified the behaviour beliefs cause as behaviour that tends to realize desires, and this bit of inter-defining looks rather like circularity. However, the characterization can be improved in a way which makes it clear that this is not so - and, as an added bonus, also makes clear just what would be involved in the denial of the existence of propositional attitudes. The secret is to take the general account of the definition of theoretical terms provided by David Lewis and apply it to the propositional attitudes.

Roughly, the idea is to collect together a class of sentences involving the verbs of propositional attitude and treat this class as a simultaneous definition of the propositional attitudes. The same idea can be applied to other webs of theoretical terms. A very simple case is the definition of a wife as someone with a husband, and a husband as someone with a wife. We remove the air of circularity when we note that we could spell matters out as follows: x is a wife if there are two people, one male and one female, they are married to each other, and x is the female one. Consider, for a more complex example, the artworld and its denizens. Artists are producers of art works. Audiences are viewers or consumers of art works. Critics are critical assessors of art works. Galleries are places for the exhibition of art works. And art works are objects produced by artists for consumption by audiences, assessment by critics and display in galleries. Taken together, these sentences - or some more elaborate version thereof - can be taken to provide a simultaneous definition of the key terms: artist, audience, critic, gallery, artwork. And there is no (vicious) circularity involved because the other terms which feature in the various sentences - 'production', 'consumption', 'display', 'object', 'view', and so on - are (assumed to be) independently understood. These terms in effect specify the key interconnections, and the definition of artist, work of art, and so on that is delivered amounts to saying that an artist is anyone who is appropriately interconnected.

The sentences which are collected together are either *prima facie* analytic truths - that is, conditionals and biconditionals which encode the inferential practices constitutive of possession of the concepts under analysis - or else claims involving the concepts to be analysed which are regarded as *prima facie* common knowledge by those, or most of those, who possess the concepts. Under the former account, there is no guarantee that the sentences will be immediately available to those who possess the concepts; however, it might be hoped that reasonable and reflective people can be brought to assent to them when the sentences are drawn to their attention. Under the latter account, there is no requirement that the sentences encode inferential practices constitutive of possession of the

concepts under analysis: for all kinds of synthetic and a posteriori claims may be selected. However, the following constraint is to be observed: it should be commonly held to be a condition on attribution of possession of the concepts which are up for analysis to a subject that they do not reject, or fail to accept, too many of the chosen sentences.

On either account, one might doubt that there are sentences of the kind required by the analysis. So, for example, it might be said that it is notoriously difficult to provide exceptionless generalizations about the connections between beliefs, desires, intentions, actions, and so on. However, it is important to note that there is nothing in the account which requires that the platitudes in question must be couched as exceptionless generalisations. Indeed, a plausible thought is that most of the platitudes will instead make claims about what is normally the case - that is about what happens when all other things are equal, or when conditions are normal. So, for example, on either account, there are sentences like these: a system of beliefs and desires tends to cause behaviour that serves the subject's desires according to that subject's beliefs; beliefs typically change under the impact of sensory evidence; desires aim to have the world fit them; beliefs aim to fit the world, and so on.

The final point to note is that our definition tells us that the propositional attitudes are whatever it is that satisfies the chosen set of sentences and so stand in the specified relations. In the simplest case, there will be a unique natural collection of candidates for the propositional attitudes. In that case, the sentences provide an explicit definition of the propositional attitudes. But things may not be so simple: perhaps there is nothing which satisfies the sentences exactly; and perhaps the only thing which satisfies the sentences is extremely unnatural, that is, gerrymandered. In these cases, we shall look for the best near-satisfier, where the criteria which guide our search are (1) considerations of naturalness; and (2) satisfaction of as many as possible of the more central platitudes. Put another way: application of a procedure which aims at achieving reflective equilibrium may lead to the rejection of some of the sentences, and perhaps to the elevation of further sentences to membership of the favoured class. Moreover, the procedure which aims at achieving reflective equilibrium is guided by judgments about (1) the naturalness of the resulting concept; and (2) the centrality or revisability of the sentences which are incorporated or rejected.

Of course, there is nothing in the account of this procedure which guarantees that there will be a point of reflective equilibrium: and, in the case in which there is no point of reflective equilibrium, we will need to give an error theory of the propositional attitudes - that is we will hold either that there were no coherent concepts characterized by our pre-reflective usage, or else that there simply is nothing to which our concepts apply.

6 Scepticism about propositional attitudes

There are various sceptical doubts which philosophers have had about propositional attitudes. Only some of these doubts will be mentioned here.

Some philosophers - W.V.O. Quine, Paul Churchland, Stephen Stich, for example - deny that there are any propositional attitudes - they deny that there are any beliefs, desires, intentions and the like. Sometimes, the source for this denial is philosophy of language (dislike of intensional contexts, and so on); other times, the source is metaphysics (failure to find a physical structure with which the states in question can be identified, for example). Sometimes, the view is that propositional attitudes are convenient fictions - useful instruments for making predictions of behaviour, but not suited to the business of serious science. On this view, talk of propositional attitudes has a second grade, merely instrumental, status, but there is no reason why we should not continue with it for purposes of prediction and explanation - a common comparison is with talk about the average family. Other philosophers argue that talk of propositional attitudes is intended as first grade scientific theorizing, but that it just happens to be seriously astray. On this view, talk of propositional attitudes is like talk of phlogiston: talk which sought to capture an important feature of reality and to be part of serious science, but failed (see Eliminativism).

Among philosophers who accept that there are propositional attitudes - that is, among those philosophers who accept that some propositional attitude ascriptions are strictly and literally true - there are philosophers who are sceptical of the relational analysis of those attitudes. Thus, some philosophers suppose that propositional attitudes should be given an adverbial analysis according to which, for example, one's belief that p is a matter of one's believing p-ly (see Adverbial theory of mental states). They deny that the attitudes are relations to propositions. Sometimes, this denial is fuelled by worries about the candidate entities for those contents, the sets of worlds,

structured set-theoretic entities, and the like we discussed earlier. Theorists who deny the relational view of the attitudes typically insist that a sentence like 'I believe that snow is white' is best written, from a logical point of view, as 'I snow-is-white believe', somewhat as 'I have a limp' is best written 'I limp'.

Yet other philosophers - Stephen Schiffer, for example - despair of the project of giving any philosophical account (theory, analysis) of propositional attitudes, while nonetheless supposing that our practice of making propositional attitude ascriptions is perfectly in order as it stands. The essence of their position is a denial of the need to give an account in anything like the traditional sense of the propositional attitudes.

Theorizing about propositional attitudes is a very difficult task. This brief summary only indicates a few of the controversial questions which such theorizing is bound to confront.

See also: Action; Folk psychology; Intention; Desire; Propositional attitude statements; Communication and intention

GRAHAM OPPY

References and further reading

Braddon-Mitchell, D. and Jackson, F. (1996) *The Philosophy of Mind and Cognition*, Oxford: Blackwell. (Excellent, state-of-the-art textbook for philosophy of mind.)

Churchland, P. (1981) 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy* 78: 67-90.(Well-known defence of a style of eliminativism about the propositional attitudes that sees propositional attitude talk as failed science.)

Davidson, D. (1968) 'On Saying That', *Synthèse* 19: 130-46.(Treatment of the contents of propositional attitude ascriptions in terms of sentences in English, but not as sentences of English.)

Fodor, J. (1975) *The Language Of Thought*, New York: Crowell(Defence of the hypothesis that there is a language of thought.)

Frege, G. (1917) 'Thoughts' in P. Geach (ed.) *Logical Investigations*, Oxford: Blackwell, 1977.(Famous defence of the view that the contents of propositional attitudes are abstract structured entities.)

Grice, P. (1957) 'Meaning', *Philosophical Review* 66: 377-88.(A classic. Includes famous analysis of meaning in terms of propositional attitudes.)

Harman, G. (1972) *Thought*, Princeton, NJ: Princeton University Press.(Early, clear, exposition of a functionalist account of the mind and, in particular, the propositional attitudes.)

Kripke, S. (1982) *Wittgenstein on Rules and Private Language*, Oxford: Blackwell.(Classic text which deals, *inter alia*, with problems about the normativity of propositional attitude contents.)

Lewis, D. (1970) 'How To Define Theoretical Terms', *Journal of Philosophy* 67: 427-46.(Masterly treatment of the definition of theoretical terms. Much better - but much harder - than the outline in §5!)

Lewis, D. (1994) 'Reduction of Mind', in S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*, Oxford: Blackwell, 412-31.(Superb exposition of Lewis's philosophy of mind. Assumes some familiarity with the prior literature.)

Loar, B. (1981) *Mind and Meaning*, Cambridge: Cambridge University Press.(Difficult, but very good, discussion of propositional attitudes, and their importance for philosophy of mind and philosophy of meaning.)

Pessin, A. (ed.) (1995) *The Twin Earth Chronicles*, London: Pergamon Press.(Collection of articles about Twin Earth, broad and narrow content, and so on. Good reading list.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press and John Wiley.(Famous and elegant text which includes defence of a kind of eliminativism about propositional attitudes.)

Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.(Defence of treatment of propositions as structured, set-theoretic entities.)

Salmon, N. and Soames, S. (eds) (1988) *Propositions and Attitudes*, Oxford: Oxford University Press.(Excellent collection of papers about the semantics of propositional attitude ascriptions. Good reading list.)

Schiffer, S. (1972) *Meaning*, Oxford: Clarendon Press.(Defence of the view that propositional attitudes - in particular, belief and intention - are the key to meaning. Not easy going.)

(1987) *Remnants of Meaning*, Cambridge, MA: MIT Press.(Retraction of the view that the propositional attitudes - in particular, belief and intention - are the key to meaning, and of much else besides. Still not easy going.)

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press.(Defence of the view that the contents of propositional attitudes are sets of possible worlds. Includes best available treatment of the objection that any such treatment makes the objects of the attitudes too coarse-grained.)

Stich, S. (1983) *From Folk Psychology to Cognitive Science: The Case Against Belief*, Cambridge, MA: MIT Press.(Case for eliminativism about belief, and other propositional attitudes. Interesting, and reasonably accessible.)

Propositions, sentences and statements

A sentence is a string of words formed according to the syntactic rules of a language. But a sentence has semantic as well as syntactic properties: the words and the whole sentence have meaning. Philosophers have tended to focus on the semantic properties of indicative sentences, in particular on their being true or false. They have called the meanings of such sentences 'propositions', and have tied the notion of proposition to the truth-conditions of the associated sentence.

The term 'proposition' is sometimes assimilated to the sentence itself; sometimes to the linguistic meaning of a sentence; sometimes to 'what is said'; sometimes to the contents of beliefs or other 'propositional' attitudes. But however propositions are defined, they must have two features: the capacity to be true or false; and compositional structure (being composed of elements which determine their semantic properties).

One reason for distinguishing a sentence from 'what the sentence says' is that a sentence may be meaningless, and hence say nothing, yet still be a sentence. But perhaps the main reason is that two people, *A* and *B*, may utter the same sentence, for example, 'I am hot', and say the same thing in one sense but not in another. The sense in which they say the same thing is that they use the same words with the same linguistic 'meaning'. The sense in which they say something different is that they put the same words to different 'uses': *A* uses 'I' to refer to *A*, while *B* uses it to refer to *B*. Hence what *A* says may be true although what *B* says is false. If what is said can be true in one case and false in another, *A* and *B* have not made the same 'statement'. On the other hand, if *B* utters instead 'You are hot' when *A* utters 'I am hot', they put different sentences to the same use and make the same statement. On this view (see Strawson 1952), we must distinguish the 'sentence', the 'use' of the sentence, and the 'statement' made by using a sentence in a context of utterance. According to Strawson, it is not sentences but statements that are true or false.

Logicians usually abstract from the context of utterance of sentences in actual communication and talk of the propositions expressed as abstract entities. The main modern proponent of this conception is Frege. For Frege (1918), a proposition is a 'thought', which is both the cognitive meaning expressed by a sentence and the content of a propositional attitude such as belief or desire; thoughts are the 'senses' of sentences. Thoughts are distinguished according to the following principle: if it is possible rationally to believe that *p* and not to believe that *q* then the thought that *p* and the thought that *q* are distinct (see Frege, G. §§3-4; Sense and reference §2).

Under a different, but related, conception, proposed by Carnap (1947), the proposition expressed by a sentence *S* is the set of possible worlds in which *S* is true (see Semantics, possible worlds §9-10). This view violates the compositional structure requirement, since such propositions are not grasped by grasping their components (see Compositionality). It also seems unable to differentiate between some distinct propositions. For example, there would only be one necessarily false proposition, since there is only one empty set of worlds, but intuitively there are many different necessarily false propositions.

An alternative view of propositions as entities is the (neo-)Russellian view that they are collections of actual entities making up 'facts' or 'states of affairs' (see Facts §1). Thus the proposition expressed by 'Socrates is mortal' is the ordered pair $\langle \text{Socrates, being mortal} \rangle x$, composed of the individual Socrates and the property of being mortal. Three questions for this view are: (1) Are the contents of false sentences negative facts, and are there such facts? (2) What is the criterion of identity for facts? (3) How can propositions as facts be the contents of propositional attitudes? The last of these can be seen as a version of the problem which led Frege to postulate senses: for the fact that $a = b$ must be the same fact as $a = a$, although someone who believes that $a = a$ does not thereby believe that $a = b$, and hence the propositions are not the same (see Sense and reference §1). Fregeans conclude that this shows that we cannot dispense with the notion of sense. If Russellians deny this, they must complicate their account of facts or their account of propositional attitudes (see Propositional attitude statements).

Whether one defines the intuitive notion of 'what is said' as a context-independent entity or by recourse to the notion of a statement or proposition expressed by an utterance in a particular context, one must say what it means for two sentences to express 'the same proposition' or make 'the same statement'. Both phrases rely on the notions of meaning and synonymy, criticized by Quine (1968; see Quine, W.V. §8). But if we dispense altogether with

propositions, statements or any other notion of the content of what is said, and choose instead to take sentences as truth bearers, we face two problems. The first is that sentences are unlikely candidates for the role of contents of propositional attitudes: if 'I believe I am stupid' introduces a relation between me and the English sentence 'I am stupid', it should be translated into German (say) as 'Ich glaube "I am stupid"', although the correct translation is 'Ich glaube daß ich dumm bin', which says nothing about any English sentence (Church 1950). The second problem is that, in spite of our ontological scruples about admitting propositions as entities, we still need such a notion to express the content of what is said or believed, and to account for the intentionality of thought in general. In that respect, however vague or ill-defined the notion of a proposition, it cannot be dispensed with.

See also: Intensional entities; Logical and mathematical terms, glossary of

PASCAL ENGEL

References and further reading

Blackburn, S. (1975) 'The Identity of Propositions', in S. Blackburn (ed.) *Meaning, Reference and Necessity*, Cambridge: Cambridge University Press.(A good statement of a deflationary alternative to the ontological conception of the problem.)

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press, 2nd edn, 1956.(The first modern formulation of the possible worlds account.)

Cartwright, R. (1962) 'Propositions', in R.J. Butler (ed.) *Analytical Philosophy, First Series*, Oxford: Blackwell, 81-103.(A classical discussion of the main issues.)

Church, A. (1950) 'On Carnap's Analysis of Statements of Assertion and Belief', *Analysis* 10: 97-9.(States his translation argument against sentences as objects of belief.)

Frege, G. (1918) 'Der Gedanke: eine logische Untersuchung', *Beiträge zur Philosophie des deutschen Idealismus* 1: 58-77; trans. P.T. Geach and R. Stoothoff, 'Thoughts', in *Collected Papers*, ed. B.F. McGuiness, Oxford: Blackwell, 1984, 351-72.(The classic statement of propositions as 'thoughts' or abstract entities.)

Geach, P.T. (1965) 'Assertion', *Philosophical Review* 74(4); repr. in *Logic Matters*, Oxford: Blackwell, 1968.(A criticism of the notion of statement.)

Nuchelmans, G. (1973) *Theories of the Proposition*, Amsterdam: North Holland.(The best historical account, although somewhat doxographic.)

Quine, W.V. (1968) 'Propositional Objects', *Critica* 5: 3-22; repr. in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 1969.(Includes a summary of Quine's misgivings about propositions as entities.)

Russell, B.A.W. (1918) 'The Philosophy of Logical Atomism', in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Routledge, 1992.(Treats propositions as states of affairs in the world.)

Salmon, N. and Soames, S. (eds) (1988) *Propositions and Attitudes*, Oxford: Oxford University Press.(A good anthology on the semantics of propositional attitudes, focused on the Russellian view.)

Stalnaker, R. (1976) 'Propositions', in A.F. MacKay and D.D. Merrill (eds) *Issues in the Philosophy of Language*, New Haven, CT: Yale University Press, 79-92.(A good account of the possible worlds view.)

Strawson, P.F. (1952) 'On Referring', *Mind* 59: 320-44; repr. in *Logico-Linguistic Papers*, London: Methuen, 1971.(Puts the classical arguments in favour of the notion of statement.)

Questions

Some theorists hold that a question is an interrogative sentence; others that a question is what is meant or expressed by an interrogative sentence. Most theorists hold that each question has two or more answers, and that the point of asking a question is to have the respondent reply with one of the answers. Most hold that each question has an assertive core or presupposition that is implied by each of the answers; if it is false, then no answer is true, so we say that the question commits the fallacy of many questions and we regard the negation of the presupposition as a corrective reply to the question (it corrects the question).

For example, consider the question 'Has Adam stopped sinning?' Its answers are 'Adam has stopped sinning' and 'Adam has not stopped sinning'. It presupposes 'Adam has sinned'; thus 'Adam has not sinned' is a corrective reply. The 'safe' way to ask this question is via the conditional 'If Adam has sinned, then has Adam stopped sinning?'

We can construct formal systems for asking whether *and* which *questions in an effective way. Other types of question (for example,* who *and* why*) are still problematic. It can be proved that some questions are reducible to others, some questions raise others, and some systems for the logic of questions can never be complete in certain ways.*

1 Concepts and theories

Questions have been discussed since Aristotle, although formal, systematic study did not begin until the 1950s. Theorists disagree on basic concepts, and there are at least three approaches to theorizing about questions. In one, the essence of a question is its set of answers, in another the essence is the questioner's intentions, and in another the essence is an objective intensional entity.

The 'reduction-to-answers' view holds that every question has direct answers, these direct answers are statements, and to know what counts as a direct answer to the question is to understand the question. A question presupposes any statement that is implied by every direct answer. A complete answer is a statement that implies a direct answer; a partial answer is a statement that is implied by a direct answer. The reduction-to-answers view is exemplified in most of the systems noted in §2 below.

The 'reduction-to-*intentions*' view holds that to know the questioner's intentions is to understand the question (see §4 below). The 'reduction-to-*intensions*' view holds that questions are intensional entities that are not relative to language and might or might not be expressible by sentences in a language (see §5 below and Intensional entities).

2 Effective systems

For many purposes (for example, conversation with strangers, courtroom procedure, automated information retrieval systems), it is desirable to have systems for question and answer that are effective in the following sense: for any expression of the given language, we can effectively tell whether it is an interrogative sentence and, if it is one, tell what question it expresses. In the reduction-to-answers view this implies that, given an interrogative sentence, we can effectively tell what counts as a direct answer to the question that is expressed.

Beginning with G. Stahl and T. Kubiński in the 1950s, logicians have constructed systems that provide for *whether* and *which* questions and that satisfy the effectiveness condition. The system of Belnap and Steel (1976) is exemplary. In it, *whether* and *which* questions are expressed by interrogatives of the form *?RS*. Here the subject part *S* indicates a set of statements, which are the alternatives presented by the question. (For *whether* questions *S* is a finite list of statements; for *which* questions *S* is a formula indicating all the statements that result when names are substituted for the queriables (the free variables) in the formula.) Given *S*, the request part *R* indicates how many of the true alternatives are to be put into each direct answer and indicates whether the answer is to claim that it contains all of the true alternatives and (in the case of *which* questions) claim that the various names given in the answer denote distinct entities. The system provides for *who*, *when*, *where*, *why*, *what* questions - if these are construed as asking which person, which time, which place, which reason, which thing. It allows the questioner to call for a complete list or just some examples or an answer of the form at-least-*m*-but-at-most-*n*.

3 Some problems

Systems such as those noted above are good models for some question-asking situations but not for others. First consider 'What is in the box?' and 'Some of John's things are in the box'. If the latter is a complete answer, as some theorists say, then this *what* question cannot be construed as a *which things* question, and the system must be altered or extended to provide for it. Similar examples exist for *who* and other so-called 'wh- questions'.

Why questions pose special problems. Consider 'Why did this apple fall?'. We may construe this as 'What is a good explanation of why this apple fell?', but we will not know what counts as an answer until we know what counts as a good explanation. S. Bromberger (1992) has proposed that some types of *why* question can be adequately answered by citing (1) a general rule, (2) an abnormic law - that is, a law that specifies exceptions to the rule, and (3) some exception mentioned by the law. For example: 'An apple remains on the tree unless it becomes ripe or a wind blows or the tree dies, and in this case the wind blew.' It is reasonable to require that the abnormic law be empirical and true, but then (because we cannot always recognize truth) we cannot always recognize the admissible direct answers.

Effective systems have some intrinsic limitations. Suppose that the interrogatives of a system can be effectively recognized and hence listed by some algorithm and that, for each interrogative, its direct answers can be recognized and hence listed. Then it can be shown by Cantor's diagonal method that the system is incomplete in the sense that there is a set of sentences that might be the answers to some question but cannot be the answers for any interrogative in the given list (Harrah 1984).

4 Epistemic analysis of questions

L. Åqvist (1965) proposed that questions are requests concerning epistemic states, expressible by imperatives such as 'Let it be the case that I know that *A* or I know that *B*' or 'Let it be the case that, for some *x*, I know that *Px*'. In this approach there may be a variety of imperative operators, and reference made to different knowers, so that, for example, a teacher can put questions to a class in the form 'Bring it about that I know that you know…' J. Hintikka (1983) has developed this approach further to elaborate its pragmatic dimension. Hintikka has extended his game-theoretic semantics to provide a semantics for epistemically construed questions and rules for conducting various types of rational question-and-answer dialogue (see Semantics, game-theoretic).

In this approach every question has a presupposition (in the first example above: '*A* or *B*'), a desideratum ('I know that *A* or I know that *B*') and direct answers (*A*, *B*). If an answer meets certain conditions of relevance and conclusiveness, guaranteeing that the desideratum is satisfied, then it is a conclusive answer. The conditions vary with the type of question. In the case of a *which* question, quantified variables should range over entities with which the questioner is acquainted, and the answer should specify a noun phrase only if the questioner knows what that phrase denotes. There is no guarantee that the respondent can effectively tell from the imperative sentence what the range of the variables is supposed to be, or what noun phrases will suffice, so, in this approach, the concept of conclusive answer in general is not effective.

5 Intensional analysis of questions

P. Tichý (1978) proposed that a question is a function defined on possible worlds. Propositions, individual concepts and properties are such functions, and they are the common types of question. For any given world the value of such a function is a truth-value, an individual or a set of individuals, respectively. In general the right answer to a question is the value of the function in the actual world.

This idea can be developed on its own; it can also be combined fruitfully with the ideas of Richard Montague. In Montague's analysis of declaratives, meanings are higher-order intensional entities. These are assigned to linguistic expressions according to Frege's principle that the meaning of a syntactical compound is the corresponding semantic compound of the meanings of the parts of the syntactical compound. Some linguists and logicians have developed analyses of this sort that treat both declaratives and interrogatives. In most, an interrogative expresses a question and denotes its true direct answers (see Frege, G.).

6 Erotetic logic

In a general sense erotetic logic is the logic of utterance and reply (Harrah 1984). In the strict sense it is the logic

of question and answer.

For most systems, especially those noted in §2 above, an important concept is that of containment. One question contains (or covers, or obviates) another if and only if every direct answer to the first implies some direct answer to the second, and two questions are equivalent if and only if each contains the other.

For analysis of the raising of questions, the implying of questions and arguing to questions, Wiśniewski (1995) has defined some useful concepts. For example: A set of declarative sentences evokes a question if and only if the set implies that the question has a true direct answer but does not imply the truth of any particular direct answer. Also, one question implies another if and only if each direct answer to the first implies that some direct answer to the second is true and for each direct answer to the second there is a proper subset (that is, a small selection) of direct answers to the first such that the given direct answer to the second implies that a true direct answer to the first is contained in that subset (so each true answer to the second helps us find a true answer to the first).

Another of Wiśniewski's concepts is that one question is reducible to a set of other questions (call it the reducing set) if and only if each direct answer to the given question implies that all the questions in the reducing set have true direct answers; and each collection of sentences that contains exactly one direct answer to each question in the reducing set implies some direct answer to the given question; and no question in the reducing set has more direct answers than the given question does. One result is that, if a question is 'safe' (that is, in every interpretation of the language the question has a true direct answer), then the question is reducible to some set of simple yes-no questions.

<div align="right">DAVID HARRAH</div>

References and further reading

All the works cited contain some technical material. All contain useful bibliographies.

Åqvist, L. (1965) *A New Approach to the Logical Theory of Interrogatives*, Uppsala: Almqvist & Wiksell. (Referred to in §4. Outlines an approach based on commands about knowledge. Technical and difficult.)

Belnap, N.D., Jr and Steel, T.B., Jr (1976) *The Logic of Questions and Answers*, New Haven, CT: Yale University Press.(Referred to in §2. Presents many useful concepts.)

Bromberger, S. (1992) *On What We Know We Don't Know: Explanation, Theory, Linguistics, and How Questions Shape Them*, Chicago, IL: University of Chicago Press.(Referred to in §3. Good discussion of *why* questions.)

Groenendijk, J. and Stokhof, M. (1984) *Studies on the Semantics of Questions and the Pragmatics of Answers*, Amsterdam: Academisch Proefschrift.(Gives an intensional analysis with a pragmatic dimension.)

Harrah, D. (1984) 'The Logic of Questions', in D. Gabbay and F. Guenthner (eds) *Handbook of Philosophical Logic, Vol. II: Extensions of Classical Logic*, Dordrecht: Reidel, 715-64.(Referred to in §§3, 6. Introduction, some history and a survey of various theories.)

Harrah, D. (1985) 'A Logic of Message and Reply', *Synthèse* 65: 275-94.(§3 presents a general erotetic logic.)

Hintikka, J. (1983) 'New Foundations for a Theory of Questions and Answers', in Kiefer, F. (ed.) *Questions and Answers*, Dordrecht: Reidel, 1983, 159-90.(Referred to in §4. Outlines an approach based on a modification of Åqvist (1965).)

Hintikka, J. (ed.) (1988) 'Knowledge-Seeking by Questioning', *Synthèse* 74: 1-262.(Papers by various authors, mainly on Hintikka's approach.)

Hintikka, J. (1992) 'The Interrogative Model of Inquiry as a General Theory of Argumentation', *Communication and Cognition* 25: 221-42.(Brief introduction to Hintikka-style systems.)

Kiefer, F. (ed.) (1983) *Questions and Answers*, Dordrecht: Reidel.(Papers presenting several approaches and discussing many aspects.)

Tichý, P. (1978) 'Questions, Answers, and Logic', *American Philosophical Quarterly* 15: 275-84.(Referred to in §5. Outlines an approach based on a theory of possible worlds.)

Wiśniewski, A. (1995) *The Posing of Questions: Logical Foundations of Erotetic Inferences*, Dordrecht: Kluwer. (Concerned mainly with raising and implying, as described in §6; technical and rigorous in some chapters but very clear throughout; has useful expositions, surveys and bibliography, and can serve as an introduction to the entire field.)

Radical translation and radical interpretation

Radical translation *is the setting of a thought experiment conceived by W.V. Quine in the late 1950s. In that setting a linguist undertakes to translate into English some hitherto unknown language - one which is neither historically nor culturally linked to any known language. It is further supposed that the linguist has no access to bilinguals versed in the two languages, English and (what Quine called) 'Jungle'. Thus, the only empirical data the linguist has to go on in constructing a 'Jungle-to-English' translation manual are instances of the native speakers' behaviour in publicly recognizable circumstances. Reflecting upon the fragmentary nature of these data, Quine draws the following conclusions:*

(1) It is very likely that the theoretical sentences of 'Jungle' can be translated as wholes into English in incompatible yet equally acceptable ways. In other words, translation of theoretical sentences is indeterminate. On the assumption that a sentence and its translation share the same meaning, the import of indeterminacy of translation is indeterminacy of meaning: the meanings of theoretical sentences of natural languages are not fixed by empirical data. The fact is, the radical translator is bound to impose about as much meaning as they discover. This result (together with the dictum 'no entity without identity') undermines the idea that propositions are meanings of sentences.
(2) Neither the question of which 'Jungle' expressions are to count as terms nor the question of what object(s), if any, a 'Jungle' term refers to can be answered by appealing merely to the empirical data. In short, the empirical data do not fix reference.

The idea of radical interpretation *was developed by Donald Davidson in the 1960s and 1970s as a modification and extension of Quine's idea of radical translation. Quine is concerned with the extent to which empirical data determine the meanings of sentences of a natural language. In the setting of radical interpretation, Davidson is concerned with a different question, the question of what a person could know that would enable them to interpret another's language. For example, what could one know that would enable the interpretation of the German sentence 'Es regnet' as meaning that it is raining? The knowledge required for interpretation differs from the knowledge required for translation, for one could know that 'Es regnet' is translated as 'Il pleut' without knowing the meaning (the interpretation) of either sentence. Beginning with the knowledge that the native speaker holds certain sentences true when in certain publicly recognizable circumstances, Davidson's radical interpreter strives to understand the meanings of those sentences. Davidson argues that this scenario reveals that interpretation centres on one's having knowledge comparable to an empirically verified, finitely based, recursive specification of the truth-conditions for an infinity of sentences - a Tarski-like truth theory. Thus, Quine's radical translation and Davidson's radical interpretation should not be regarded as competitors, for although the methodologies employed in the two contexts are similar, the two contexts are designed to answer different questions. Moreover, interpretation is broader than translation; sentences that cannot be translated can still be interpreted.*

1 Radical translation

Suppose that children, in the normal course of events, acquire their native tongue (for example, English) by observing their parents and others talking in publicly observable circumstances. A corollary of this common sense supposition is that whatever there is to linguistic meaning can be manifested in behaviour in publicly observable circumstances. However, one might ask, just how far do empirical data go toward fixing the meanings of the sentences being spoken? Quine's thought experiment of radical translation is designed to answer this question.

In the context of radical translation, the field linguist sets out to construct a manual for translating the newly discovered language of 'Jungle' into English. Of course, the linguist has a prior knowledge of English, and will exploit that knowledge in constructing a 'Jungle-to-English' translation manual. Obviously, a child learning a first language has no prior language to exploit. So, there is this fundamental difference in psychological readiness to learn 'Jungle' between the linguist and the child of native-speakers of the language. However, this difference does not affect the central fact that the ultimate empirical data available to both linguist and child are the same: facts about behaviour. The poverty of these ultimate data is what the context of radical translation is meant to bring out.

2 Indeterminacy of translation

The primary reference for indeterminacy of translation is 'Translation and Meaning', the second chapter of Quine's *Word and Object* (1960). According to that account, the linguist first compiles a list of phonemes of the native-speakers' language (which Quine dubbed 'Jungle'), and then they settle on some expressions used by native speakers for assent and dissent.

A rabbit scurries by; the native speaker utters 'Gavagai'. Having noticed the rabbit, and that the native speaker noticed it too, and suspecting that the rabbit probably prompted the native speaker's utterance of 'Gavagai', the linguist tentatively enters into their 'Jungle-to-English' translation manual that the native speaker's 'Gavagai' might well be translated as 'Lo, a rabbit'. On later occasions, some rabbited, some rabbitless, the linguist will volunteer 'Gavagai?' with the intention of eliciting the native speaker's assent, dissent, or neither. In this way, the linguist can assess whatever inductive support there may be for their hypothesis that 'Gavagai' translates as 'Lo, a rabbit'.

The distal stimulus, in this case a rabbit, serves as the criterion for the linguist's inductive hypothesis that 'Gavagai' can be translated as 'Lo, a rabbit'. However, what a sentence (even a one word sentence) and its translation share is the same meaning. In the case at hand, is the rabbit the meaning of 'Gavagai'? No; that would be to confuse meaning with reference. What, then, is its meaning? Quine's answer to this question (what he calls 'stimulus meaning') is given in terms of the proximal stimulus, namely, patterns of a subject's activated nerve endings. Affording the native speaker a global pattern of non-verbal stimulation for a specific duration together with the linguist's query 'Gavagai?' will elicit from the native speaker either assent, dissent, or neither. Patterns of stimulation which would prompt assent and patterns which would prompt dissent belong, respectively, to the affirmative stimulus meaning and negative stimulus meaning for that sentence and native speaker at that time. Stimulus meaning *simpliciter* is the ordered pair of the two. Patterns which would prompt neither assent nor dissent are irrelevant.

Given stimulus meaning, one might say that what 'Gavagai' and its translation 'Lo, a rabbit' share is the same stimulus meaning. However, this too would be wrong, for stimulus meaning (as just defined) is private: the native speaker's nerves are distinct from the linguist's nerves. It would be more accurate, but still problematic, to say that the native speaker's stimulus meaning for 'Gavagai' and the linguist's for 'Lo, a rabbit' are approximately the same. This is Quine's approach in *Word and Object* (although he subsequently altered this account, as we shall see).

In the context of radical translation, it is stipulated that there are no bilinguals for the linguist to consult. So, the linguist must go about compiling the translation manual by observing the behaviour of native speakers, verbal and otherwise, and querying them with sentences for assent/dissent under various publicly observable circumstances. Thus, the criteria of translation remain keyed to distal stimuli, though the definition of stimulus meaning remains keyed to proximal stimuli. After all, one cannot plausibly expect the linguist to know anything about the native speakers' or their own global patterns of activated nerve endings.

Given the behavioural limits of the empirical data of radical translation, Quine maintains in *Word and Object* that (1) observation sentences, like 'Gavagai', can be determinately translated; (2) truth functions (for example, 'and', 'or', 'not') can be determinately translated; (3) stimulus-analytic and stimulus-contradictory sentences can be identified, but not translated; and (4) questions regarding the intrasubjective stimulus synonymy of occasion sentences can be settled, if raised. (A stimulus-analytic sentence is one that a person would assent to after every stimulation: 'There have been black dogs', for example. A stimulus-contradictory sentence is one that a person would dissent from after every stimulation. The occasion sentences 'Bachelor' and 'Unmarried man' could count as intrasubjectively stimulus synonymous for a person, if they have the same stimulus meaning for him.)

However, if the linguist were to become bilingual (for example, if they forgot about translation and simply acquired 'Jungle' as a native speaker might, or nearly so), then according to Quine (1′) they could determinately translate all occasion sentences, not just observation sentences, and (4) above becomes superfluous. (Non-observational occasion sentences are those that require a new prompting stimulus each time they are queried, but lack the community-wide agreement characteristic of observation sentences, such as 'There goes John's old tutor'.) This leaves the linguist of *Word and Object* with determinate translations for (1′), (2) and (3).

However, in *Roots of Reference* (1974), Quine explains that the truth functions cannot be determinately translated

after all. One barrier concerns conjunction ('and'): if each of a pair of sentences commands neither assent nor dissent, their conjunction can sometimes command dissent and sometimes neither dissent nor assent. If (to use Quine's examples) 'It's a mouse' and 'It's a chipmunk' are neither affirmed nor denied, still their conjunction will be denied. Whereas if the components are 'It's a mouse' and 'It's in the kitchen' and neither is affirmed or denied, their conjunction will be neither affirmed nor denied. An analogous barrier stands in the way of determinately translating alternation ('or').

3 Indeterminacy of translation (cont.)

In discussing the translation of observation sentences and truth functions in Word and Object, Quine makes use of Neil Wilson's principle of charity (see Wilson 1959). The central idea of the principle is that any proffered translation that construes the native speaker as holding some patently silly belief (for example, that the law of non-contradiction is false) is less likely than that the proffered translation is a bad one (see Charity, principle of §4). Thus, there is a sound methodological reason for translating a native speaker so as to construe them as holding true beliefs (true by the linguist's own standards). However, Quine augments this principle with another: maximize psychological plausibility. This latter principle permits the linguist to translate a native speaker's sentence by some patently false English sentence if, given the native speaker's 'outlandish' rites and taboos or whatever, doing so is more plausible than translating the sentence in question by some true English sentence.

Putting the matter overly schematically, the next steps the linguist takes in constructing a manual of translation are as follows: (a) segmenting newly heard sentences uttered by native speakers into short recurrent parts ('words') which these sentences share with some of the observation sentences already translated; (b) segmenting further new sentences in ways that reflect the 'words' of the previous step, so that a lexicon and grammar begin to emerge; (c) generating new sentences in the native speaker's language as recursion sets in; (d) continuing to fine tune the manual. Steps (a) to (d) make use of analytical hypotheses, hypotheses that go beyond the behavioural data: though analytical hypotheses should not contravene those data, neither are they supported by them.

In 'Indeterminacy of translation again' (1987), Quine articulates some constraints of a pragmatic nature which help to guide the linguist's conjectures in this area, but clearly the linguist's latitude for conjecture remains enormous. In fact, two linguists working independently of one another might well come up with equally successful 'Jungle-to-English' manuals which, despite being consistent with all the behavioural data, differ from one another in assigning to countless 'Jungle' sentences different English sentences as translations. Of these, countless different English sentences would not be interchangeable in English contexts. But which of these rival translation manuals is the correct one? There is no fact of the matter here; they are both correct. Such is the thesis of indeterminacy of translation.

Quine sums up the philosophical import of radical translation and of indeterminacy as follows:

> The point of my thought experiment in radical translations was philosophical: a critique of the uncritical notion of meanings and, therewith, of introspective semantics. I was concerned to expose its empirical limits. A sentence has a meaning, people thought, and another sentence is its translation if it has the same meaning. This, we see, will not do.
>
> (Quine 1987: 9)

The truth is more nearly the reverse: if some sentence is the translation of another, then it is proper to say that it has the same meaning. So, the idea of propositions as objectively valid translation relations, or as meanings of sentences, cannot be maintained. Furthermore:

> The critique of meaning leveled by my thesis of indeterminacy of translation is meant to clear away misconceptions, but the result is not nihilism. Translation remains, and is indispensable. Indeterminacy means not that there is no acceptable translation, but that there are many. A good manual of translation fits all checkpoints of verbal behavior, and what does not surface at any checkpoint can do no harm.
>
> (Quine 1987: 9)

So, the radical translator imposes about as much meaning as they discover.

Before surveying some criticisms of Quine's indeterminacy thesis, we need to clear up an ancillary matter. In

Word and Object, Quine claims that the native speaker's stimulus meaning for 'Gavagai' and the linguist's stimulus meaning for 'Lo, a rabbit' are approximately the same. But Quine realized as early as 1960 that such intersubjective comparisons of stimulus meanings are problematic, for no two people's nerve nets are isomorphic. Consequently, in *Pursuit of Truth* (1990) Quine proposes to do without intersubjective comparisons of stimulus meaning. Empathy fills the gap. A rabbit scurries by; the native speaker utters 'Gavagai'. Empathizing with the native speaker, the linguist conjectures that were they in the native speaker's place they would have uttered 'Lo, a rabbit'. Thus, the linguist forms the tentative inductive hypothesis that 'Gavagai' translates as 'Lo, a rabbit'. Appealing to empathy instead of to approximate sameness of stimulus meaning forces Quine to reformulate his definition of 'observation sentence' (see Quine [1990] 1992: 43), but he does not drop the concept of stimulus meaning from his scientific semantics. Quine leaves to the neuropsychologists the task of accounting for the mechanisms underlying humans' ability to empathize.

4 Criticism

Indeterminacy of translation has been hotly debated by philosophers ever since the publication of *Word and Object*, and it is probably fair to say that more philosophers reject what they take to be the thesis than accept it. Some of the more widespread criticisms are: (a) Quine does not prove the thesis; (b) linguistics is underdetermined, but not indeterminate; (c) the thesis is unintelligible; and (d) the evidence for actual translation is not limited to that of radical translation. Let us briefly examine each of these criticisms in turn:

(a) *In 'Word and Object' Quine does not provide a deductive proof of his thesis of indeterminacy.* This is correct, but in addition to the more or less inductive argument presented there, Quine elsewhere offers two other arguments for indeterminacy. One of these, found in 'Epistemology naturalized' (1969a), takes Peirce's verificationism and Duhem's holism as premises supporting indeterminacy:

> If we recognize with Peirce that the meaning of a sentence turns purely on what would count as evidence for its truth, and if we recognize with Duhem that theoretical sentences have their evidence not as single sentences but only as larger blocks of theory, then the indeterminacy of translation of theoretical sentences is the natural conclusion.
>
> (Quine 1969a: 80-1)

A second argument is to be found in 'On the reasons for indeterminacy of translation' (Quine 1970: 179-81). There Quine argues for indeterminacy on the grounds of underdetermination of physical theory. He argues that translation of a foreign physicist's theory which is underdetermined by observable evidence will itself be underdetermined by translation of the theory's observation sentences. Thus, the translator will have to rely on analytical hypotheses, and different systems of analytical hypotheses can result in equally good, but incompatible, translations. Thus, the extent to which a physical theory is underdetermined by observation is a measure of the extent to which its translation is indeterminate. (But see Quine 1979: 66-7 for a reservation about this way of arguing for indeterminacy.)

(b) *It is true that scientific theory is underdetermined by all possible observations, and in so far as linguistic theory is a part of scientific theory it too is underdetermined, but there is no special indeterminacy that afflicts linguistics.* Quine responds to this criticism by pointing out that once one accepts some single physical theory from among its competitors as the whole truth about what there is, still, within that theory, indeterminacy of translation can occur. The reason is that all that matters to translation are the behavioural facts, and since behavioural facts do not fix translation, no further physical facts, not even the whole truth about nature, would do so.

(c) *The thesis of indeterminacy is unintelligible, for it repudiates meanings and at the same time claims that there can be rival manuals of translation assigning to some 'Jungle' sentence English sentences which diverge in meaning. But if there are no meanings, then how can English sentences diverge in meaning? And if they can diverge in meaning, then how can there be no meanings?* The response to this criticism is to explain how two English sentences can diverge in meaning without reifying meanings. Quine responds as follows:

> A manual of Jungle-to-English translation constitutes a recursive, or inductive, definition of a *translation relation* together with a claim that it correlates sentences compatibly with the behavior of all concerned. The thesis of indeterminacy of translation is that these claims on the part of two manuals might both be true and yet the two translation relations might not be usable in alternation, from sentence to sentence, without issuing in

incoherent sequences. Or, to put it another way, the English sentences prescribed as translation of a given Jungle sentence by two rival manuals might not be interchangeable in English contexts.

<div align="right">(Quine [1990] 1992: 48)</div>

Once again, Quine's solution to the difficulty is along behaviourist lines.

(d) *There may well be indeterminacy of translation in the context of radical translation, but in the context of actual translation there is more evidence available to fix the meanings of sentences. For example, it is a fact of established usage, that is, verbal dispositions, that we translate the French sentence 'Il y a peut-être des êtres intelligents sur Mars' as the English sentence 'There may be intelligent beings on Mars'.* It is true that the context of radical translation was designed to expose the ultimate empirical data for translation, where the word 'ultimate' excludes verbal behaviour linked to systems of analytical hypotheses that over time have become customary. But the point remains, even for languages as close as French and English, that some uncustomary system of analytical hypotheses could produce different translation relations between French and English which are just as effective as the customary one. Indeed, as Quine points out, indeterminacy of translation begins with the translator's home language. English, for example, could be mapped into itself by using one 'Jungle-to-English' manual to translate English into 'Jungle', then using a rival 'Jungle-to-English' manual to translate 'Jungle' back into English.

The four criticisms dealt with above are by no means the only criticisms philosophers have levelled against Quine's thesis of indeterminacy of translation, but they are some of the more interesting and widespread ones.

5 Indeterminacy of reference

As we have seen, given the empirical data, the 'Jungle' observation sentence 'Gavagai' can be determinately translated by the English observation sentence 'Lo, a rabbit'. May we conclude from this that 'gavagai' is therefore a 'Jungle' term, and if so, that it can be determinately translated as 'rabbit'? According to Quine, the answer is 'no' on both counts. Termhood and reference are matters that can be settled only by appealing to analytical hypotheses. For example, assuming the linguist decides to construe 'gavagai' as a term, is it to be construed as an abstract singular term (for example 'rabbithood'), or as a concrete general term? If the latter, then 'undetached rabbit part' could serve just as well as 'rabbit' as a translation of 'gavagai'. After all, the same portions of space-time as are occupied by undetached rabbit parts are occupied by rabbits.

The only way to settle such questions is, according to Quine, by relying upon the 'Jungle' equivalents of English plural endings, pronouns, numerals, the 'is' of identity and its adaptations 'same' and 'other'. These constitute the cluster of interrelated grammatical particles and constructions with which the individuation of terms of divided reference in English is connected. Once a linguist has fixed these equivalences, they can put to the native speaker questions like 'Is this gavagai the same as that one?', 'Is this one gavagai or two?', and so on. Once the linguist is able to ask such questions, they can begin to determine whether to translate 'gavagai' as 'rabbit' or as 'undetached rabbit part'. However, before they can ask such questions, the linguist will have to formulate a system of analytical hypotheses in connection with some other 'Jungle' expressions, namely, those playing the role of that cluster of interrelated grammatical particles and constructions of English which govern reference. But that cluster is itself susceptible to indeterminacy of translation. Thus, if one overall system of analytical hypotheses works for translating some 'Jungle' expression into 'is the same as', perhaps some other workable but systematically different system would translate the same 'Jungle' expression as 'belongs with'. Therefore, when the linguist attempts to ask 'Is this gavagai the same as that?', they could unwittingly be asking 'Does this gavagai belong with that?'. So the native speaker's assent cannot be used to settle the reference of 'gavagai' absolutely. Consequently, even though this method of translation is the best the linguist can manage, it is not sufficient for settling absolutely the indeterminacy between translating 'gavagai' as 'rabbit' or as 'undetached rabbit part', and so on. The linguist can say only that 'gavagai' refers to rabbits relative to one manual, and to undetached rabbit parts relative to a rival manual. So long as these manuals are consistent with the behaviour of all concerned, there is no fact of the matter of what is 'really' referred to by 'gavagai'.

Quine has also argued for indeterminacy of reference independent of the indeterminacy of holophrastic translation. In what he calls his 'proxy function argument', Quine explains how:

we might reinterpret every reference to a physical object arbitrarily as a reference rather to its cosmic complement, the rest of the physical universe. The old names and predicates would be introduced by ostension as usual, but it would be deferred ostension: pointing to what was not part of the intended object. Sensory associations would carry over similarly. The word 'rabbit' would now denote not each rabbit but the cosmic complement of each, and the predicate 'furry' would now denote not each furry thing but the cosmic complement of each. Saying that rabbits are furry would thus be reinterpreted as saying that complements-of-rabbits are complements-of-furry things, with 'complements-of-rabbits' and 'complements-of-furry' seen as atomic predicates. The two sentences are obviously equivalent.

(Quine 1995: 71)

Quine goes on to explain that 'cosmic complement of' expresses what he calls a 'proxy' function: a one-to-one reinterpretation of objective reference. Such proxy functions leave the truth-values of the containing sentences undisturbed.

It is a matter of reconstruing all terms and predicates as designating or denoting the proxies of what they had designated or denoted. A term that had designated an object x now designates the proxy of x, and a predicate that had denoted x now denotes the proxy of x. No big deal; we are proxying both sides of the predication, and it cancels out. We appreciated the triviality where the proxy was the cosmic complement.

(Quine 1995: 72)

Thus, unlike the conjectural status Quine accords indeterminacy of translation of theoretical sentences as wholes, indeterminacy of reference has the apodictic certainty of a proof.

6 Criticism

As with indeterminacy of translation, indeterminacy of reference (or inscrutability of reference) has been vigorously criticized. The most widespread criticism is, perhaps, that Quine's doctrine is inconsistent: it maintains that reference is and is not determinate. For example, it asserts that the native speaker's 'gavagai' is indeterminate, since it can be adequately translated by 'rabbithood', by 'undetached rabbit part', by 'rabbit', and so on, while taking the reference of these English terms to be determinate. A full Quinian response to this criticism would take us far afield, but perhaps the following two quotations will suffice:

To say that 'gavagai' denotes rabbits is to opt for a manual of translation in which 'gavagai' is translated as 'rabbit', instead of opting for any of the alternative manuals.

And does the indeterminacy or relativity extend also somehow to the home language? In 'Ontological relativity' I said it did, for the home language can be translated into itself by permutations that depart materially from the mere identity transformation, as proxy functions bear out. But if we choose as our manual of translation the identity transformation, thus taking the home language at face value, the relativity is resolved. Reference is then explicated in disquotational paradigms analogous to Tarski's truth paradigm…; thus 'rabbit' denotes rabbits, whatever *they* are, and 'Boston' designates Boston.

(Quine [1990] 1992: 52; for his essay 'Ontological relativity', see Quine 1969b)

and:

The very freedom vouchsafed us by the indeterminacy of reference allows us to *adopt* ostension as decisive for reference to observable concrete objects. We end up as we began, then, agreeing on the denotations of 'rabbit' after all: rabbits for all concerned. We may then merely differ on the deeper nature of rabbits: they are spatio temporal for some… sui generis for most. Adaptation of our usage must not, however, be allowed to obscure the lesson of proxy functions. Namely, a language-wide one-to-one reassignment of values to our variables has no effect on the truth of falsity of our statements.

(Quine 1995: 75)

We turn, now, to a consideration of radical interpretation.

7 Radical interpretation

Mary utters 'It's raining'; Tom interprets her utterance to mean that it is raining. How is this possible? What could Tom know that would enable him to interpret the sentences of a natural language? Donald Davidson argues that we would have answers to these questions - a theory of meaning - if we had a theory meeting two requirements. First, the theory must provide an interpretation for a potential infinity of utterances of a speaker or group of speakers; second, the theory must be verifiable without assuming a detailed knowledge of the speaker's propositional attitudes. The first requirement recognizes the holistic nature of linguistic understanding; the second requirement recognizes that the theory must not beg questions by assuming too much of what it is supposed to explain.

Davidson hypothesizes that a Tarski-like truth theory, suitably modified to apply to natural languages, just might meet these two conditions. To test his hypothesis, Davidson envisages a context of radical interpretation wherein an interpreter is faced with interpreting in one idiom talk in another. (For example, Kurt utters 'Es regnet' and Mary interprets his utterance to mean that it is raining.)

As Davidson points out in his essay 'Radical interpretation':

> The term 'radical interpretation' is meant to suggest strong kinship with Quine's 'radical translation'. Kinship is not identity, however, and 'interpretation' in place of 'translation' marks one of the differences: a greater emphasis on the explicitly semantical in the former.
>
> (Davidson 1984b: 126, n.1)

Or, as Quine has put the difference:

> Translation is not the field linguist's goal. His goal is to command the native language and perhaps to teach it, whether for reasons of ethnography and philology or simply to implement fluent dialogue and successful negotiation with the natives. His undertaking, broader than translation, is *interpretation*. An untranslatable sentence, such as the one about neutrinos ['Neutrinos lack mass.'] can still be interpreted, and that indeed is how we have learned it ourselves. For broadly semantic purposes, as Donald Davidson appreciates, interpretation is the thing. Translation is the narrower project, pertinent specifically to my concern over the fancied concepts of proposition and sameness of meaning.
>
> (Quine 1995: 80-1)

Thus, Quine's radical translation and Davidson's radical interpretation are devoted to different issues. Quine wants to discover the extent to which the empirical data determine meanings of sentences of a natural language; Davidson wants to discover what one could know that would enable one to understand the sentences of a natural language. Furthermore, the knowledge required for interpretation differs from the knowledge required for translation, for one could know that 'Es regnet' is translated as 'Il pleut' without knowing the interpretation (meaning) of either sentence. Davidson is fully aware of these differences, of course:

> The idea of a translation manual with appropriate empirical constraints as a device for studying problems in the philosophy of language is, of course, Quine's. This idea inspired much of my thinking on the present subject, and my proposal is in important respects very close to Quine's. Since Quine did not intend to answer the questions I have set, the claim that the method of translation is not adequate as a solution to the problem of radical interpretation is not a criticism of any doctrine of Quine's.
>
> (Davidson 1984b: 129, n.3)

What, then, is the nature of the Tarski-like theory that Davidson believes might serve as a theory of meaning?

8 Truth theory

Alfred Tarski's theory of truth is designed to apply only to artificial languages meeting certain formal requirements: extensional languages like first-order predicate logic with relations and identity. According to Davidson:

> What characterizes a theory of truth in Tarski's style is that it entails, for every sentence *s* of the object language, a sentence of the form:
>
> *s* is true (in the object language) if and only if *p*.

Instances of the form (which we shall call T-sentences) are obtained by replacing '*s*' by a canonical description of *s*, and '*p*' by a translation of *s*. The important undefined semantical notion in the theory is that of *satisfaction* which relates sentences, open or closed, to infinite sequences of objects, which may be taken to belong to the range of the variables of the object language. The axioms which are finite in number, are of two kinds: some give the conditions under which a sequence satisfies a complex sentence on the basis of the conditions of satisfaction of simpler sentences, others give the conditions under which the simplest (open) sentences are satisfied. Truth is defined for closed sentences in terms of the notion of satisfaction.

<div align="right">(Davidson 1984b: 130-1)</div>

Following Tarski's lead, Davidson's blueprint for a theory of meaning consists of axioms, rules and theorems. A finite set of axioms state satisfaction conditions for each of the object language's *semantical primitives* (basic terms and various other basic non-sentential expressions). An example might be: '*chien* applies to a thing if and only if it is a dog'. A finite set of rules allows proving T-sentences as theorems from the axioms. The resulting theorems are a potential infinity of T-sentences stating the truth-conditions for each object-language sentence. The proof of a T-sentence amounts to an analysis of how the truth or falsity of the object-language sentence depends on how it is composed from the semantical primitives. (These axioms, rules and T-sentences are, of course, stated in the language of the theory - the metalanguage.)

This general form of Davidson's theory of meaning reflects certain features of natural languages. First, since natural languages are learnable by finite beings, they must have a finite number of semantical primitives. Second, since natural languages consist of a potential infinity of sentences, the semantic features of such sentences must ultimately depend upon those of the semantical primitives. However, this does not imply that 'word' meaning is known somehow prior to sentence meaning. In fact, Davidson argues just the opposite: sentence meaning is known prior to word meaning. Words have whatever meanings they do as a result of their functioning in meaningful sentences. Third, like the axioms, the number of rules for proving T-sentences must be finite, and for the same reason. Finally, it is noteworthy that this Tarski-like theory of meaning has several obvious ontological advantages in that it makes no use of meanings as entities, nor of properties as denotatations of predicates, nor of states of affairs as denotatations of sentences.

Tarski's *Convention T* is that a theory of truth should entail T-sentences, in which descriptions of object-language sentences are linked to translations of those sentences in the metalanguage. For his own purposes, Davidson modifies Convention T in two fundamental ways. Tarski's Convention T relies on the notion of meaning (or translation) in order to explain truth. But meaning (or translation) is just the semantic notion that Davidson wants to explain, so he cannot, without begging questions, interpret Convention T in the manner of Tarski. Rather, Davidson construes Convention T as claiming that:

an acceptable theory of truth must entail, for every sentence *s* of the object language, a sentence of the form: *s* is true if and only if *p*, where '*p*' is replaced by any sentence that is true if and only if *s* is.

<div align="right">(Davidson 1984b: 134)</div>

Thus, Davidson reverses Tarski's priorities: instead of assuming meaning in order to explain truth, he assumes truth in order to explain meaning.

Furthermore, since Davidson wants to apply the theory to natural languages (something which, as Davidson reports, Tarski explicitly denies can be done), he must allow for the fact that the object language contains a wide diversity of expressions, including proper names, functional expressions, indexicals, demonstratives, and so on. Admitting indexicals and demonstratives into the object language provides for the possibility of sentences which vary in truth-value according to time and speaker. Thus, Davidson's version of a Tarskian theory must take utterances of sentences, not sentences themselves, as truth vehicles. Davidson modifies Convention T accordingly:

s is true in *L* when spoken by *x* at *t* if and only if *p*.

Recall that Davidson believes we would have an answer to the question of what a person could know that would enable them to interpret the sentences of a natural language, if we had a theory meeting two requirements. First, the theory must provide an interpretation for a potential infinity of utterances of a speaker or group of speakers - a potential infinity of T-sentences. Davidson's modified version of Tarski's theory of truth is a blueprint for

constructing such a theory. Second, since that theory is to be an empirical theory about a speaker's (or speakers') behaviour, it must be verifiable - without assuming a detailed knowledge of the speaker's (or speakers') propositional attitudes. So, how is this second requirement to be met?

This is where the notion of a radical interpreter (like Quine's radical translator) serves as a useful heuristic. By imagining the routine of a radical interpreter at work, we come to appreciate that a theory of meaning is supported by verifying its T-sentences, and we come to see what ultimate evidence there is for verifying T-sentences.

9 Truth theory (cont.)

Imagine a radical interpreter who speaks only English, does not have access to bilinguals or dictionaries, and is attempting to interpret Kurt, who speaks only German. The interpreter has no prior detailed knowledge of the meanings of Kurt's utterances nor of Kurt's beliefs. Kurt utters 'Es regnet' at a time when it is raining near him. Judging that Kurt's utterance is both linguistic and intentional, the interpreter conjectures that Kurt holds 'Es regnet' to be true. The important methodological point to note, according to Davidson, is that the interpreter can spot such utterances purely on the basis of the speaker's observable behaviour, independently of knowing anything of the utterance's meaning or the speaker's beliefs. It is those utterances that a speaker (or speakers) hold true, at a time, under publicly observable circumstances, which constitute the ultimate evidence for T-sentences.

This ultimate evidence, about what sentences a speaker holds-true, is evidence not specifically about meaning, but about a combination of meaning and belief: a speaker believes whatever it is that they mean by the sentences. However, if the interpreter knew not only that Kurt holds a certain utterance true, but also what Kurt believes at the time, then the interpreter could extract the meaning of Kurt's utterance. Similarly, if the interpreter knew both that Kurt held a certain utterance true and its meaning, then they could extract Kurt's current belief. The problem is that there is no way to discover meaning which is independent of belief, and no way to discover belief which is independent of meaning. Thus, the central problem of interpretation is simultaneously to disentangle meaning and belief. How can a radical interpreter accomplish this task?

Suppose the interpreter, to use one of Davidson's examples (1984b: 135), can formulate the following T-sentence:

(T) 'Es regnet' is true-in-German when spoken by x at time t if and only if it is raining near x at t.

The interpreter's evidence for this T-sentence is:

(E) Kurt belongs to the German speech community and Kurt holds-true 'Es regnet' on Saturday at noon and it is raining near Kurt on Saturday at noon.

The interpreter's next step is to gather generalized evidence:

(GE) $(x)(t)$ (if x belongs to the German speech community then (x holds-true 'Es regnet' at t if and only if it is raining near x at t)).

The interpreter can take (E) as evidence for (T) and for (GE), if they assume Kurt believes it is raining nearby at noon on Saturday, and if they assume that particular belief played a role in the causal chain prompting Kurt's utterance, 'Es regnet'. By using this procedure of holding belief constant as far as possible while solving for meaning, the interpreter can hope to solve the problem of the interdependence of belief and meaning.

But what justifies this procedure? After all, Kurt could be in error about it raining near him, and the interpreter could be in error about what distal stimulus is prompting Kurt's utterance and, hence, about the belief that is ascribed to Kurt.

'What justifies the procedure is the fact that disagreement and agreement alike are intelligible only against a background of massive agreement' (Davidson 1984b: 137). For example, if two people agree or disagree that what they see is a rabbit in tall grass, they must be in agreement regarding an indefinite number of other beliefs about rabbits, grass, distance, and so on. Furthermore:

> The methodological advice to interpret in a way that optimizes agreement should not be conceived as resting on a charitable assumption about human intelligence that might turn out to be false. If we cannot find a way to interpret the utterances and other behavior of a creature as revealing a set of beliefs largely consistent and true

by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything.

<div align="right">(Davidson 1984b: 137)</div>

This mandatory principle of charity indicates that the verification of ascriptions of meaning and of belief are holistic in character. So while the principle of charity (namely, 'optimize agreement') does not guarantee that Kurt's acceptance of 'Es regnet', or any other sentence considered in isolation, is to be deemed correct (in that it will be interpreted to mean something we accept), Davidson contends that the principle does guarantee that most of Kurt's acceptances are in that sense correct.

> The method is rather one of getting a best fit. We want a theory that satisfies the formal constraints on a theory of truth, and that maximizes agreement, in the sense of making Kurt (and others) right, as far as we can tell, as often as possible.

<div align="right">(Davidson 1984b: 136)</div>

Unlike Quine, who advocates the principle of charity at the level of translating 'Jungle' observation sentences like 'Gavagai' and the 'Jungle' equivalents of truth functions like 'and', 'or' and 'not', Davidson advocates the principle at every level of interpretation (see Davidson 1984b: 136, n.16). However, his principle of charity definitely involves more than the basic methodological precept 'optimize agreement':

> It is impossible to simplify the considerations that are relevant [for dealing with disagreements], for everything we know or believe about the way evidence supports belief can be put to work in deciding where the theory can best allow error, and what errors are least destructive of understanding. The methodology of interpretation is, in this respect, nothing but epistemology seen in the mirror of meaning.

<div align="right">(Davidson 1984c: 169)</div>

It is Davidson's contention that for a theory that satisfies the formal constraints on a theory of truth, and which optimally fits the evidence about utterances held true, each of its T-sentences will yield an acceptable interpretation. Moreover, Davidson believes - and for Quinian reasons - that if there is one such theory for some object language, then there are many. Thus, Davidson's theory of meaning possesses its own forms of indeterminacy and inscrutability.

10 Criticism

Davidson's theory of radical interpretation has precipitated a tremendous amount of discussion of the topic among philosophers. Two of the more general criticisms of Davidson's programme are: (a) the construction of a Tarski-like theory of meaning for an entire natural language is implausible; and (b) Davidson's rendering of Convention T is too weak to yield interpretations.

(a) *The construction of a Tarski-like theory of meaning for an entire natural language is seen to be implausible when one considers the number and variety of idioms whose truth-conditions are not fully understood; idioms like 'knows that' and 'believes that', contrary to fact conditionals, and so on.* Citing some of his own work on attributions of attitudes and performatives, as well as other philosophers' work on proper names, mass terms, 'ought', and so on, Davidson remains optimistic that many of these problems will be solved in ways consistent with his Tarski-like theory of meaning (see Davidson 1984b: 132-3).

(b) *Davidson's rendering of Convention T is too weak to yield interpretations.* Recall, for Davidson, Convention T says that *s* is true if and only if *p*, where '*p*' is replaced by any sentence that is true if and only if *s* is. Hence, we get the following T-sentence:

(1) 'John is a renate' is true if and only if John is a renate.

But, since all renates are cordates, and vice versa, we also get the following T-sentence:

(2) 'John is a renate' is true if and only if John is a cordate.

But even if (1) is an interpretation of the object-language sentence 'John is a renate', surely (2) is not. So Convention T is too weak.

Davidson's response to this problem is two-fold: he emphasizes that T-sentences are verified not individually but holistically, and he recognizes that T-sentences (and axioms) must be viewed as empirical laws. In accord with the latter point, T-sentences which yield interpretations must not only be true, they must also be capable of supporting counterfactual claims. Presumably (1) supports the counterfactual:

(1′) If 'John is a renate' were false, then John would not be a renate.

But (2) does not support the counterfactual:

(2′) If 'John is a renate' were false, then John would not be a cordate.

Davidson would explain this difference, presumably, by claiming that (1) does, but (2) does not depend ultimately on certain causal relations between speakers and the world (see Davidson 1984a: xiv).

These criticisms of Davidson's programme are by no means the only ones; many questions have been raised about his handling of indexicals, ambiguity, word-meaning versus sentence-meaning, the principle of charity, and so on.

See also: Davidson, D.; Hermeneutics; Meaning and truth; Quine, W.V.; Reference

ROGER F. GIBSON

References and further reading

Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Clarendon Press.(A collection of Davidson's essays on the philosophy of action.)

Davidson, D. (1984a) *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.(A collection containing some of Davidson's key essays on radical interpretation.)

Davidson, D. (1984b) 'Radical interpretation', in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 125-39.(One of the most important of Davidson's essays for understanding radical interpretation.)

Davidson, D. (1984c) 'Thought and talk', in *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 155-70.(One of the most important of Davidson's essays for understanding his views on the principle of charity.)

Føllesdal, D. (1988) 'Indeterminacy and mental states', in R. Barrett and R. Gibson (eds) *Perspectives on Quine*, Oxford: Blackwell, 98-109.(Føllesdal gives a clear statement of Quine's indeterminacy doctrine.)

Gibson, R. (1986) 'Translation, physics, and facts of the matter', in L. Hahn and P. Schilpp (eds) *The Philosophy of W. V. Quine*, La Salle, IL: Open Court Press, 139-54.(Gives a clear statement of the difference between indeterminacy of translation and underdetermination of physical theory; excellent bibliography of Quine's writings.)

LePore, E. (ed.) (1986) *Truth and Interpretation*, Oxford: Blackwell.(A collection of twenty-eight essays on Davidson's philosophy, including three by Davidson.)

LePore, E. and B. McLaughlin (eds) (1985) *Actions and Events*, Oxford: Blackwell.(A collection of thirty essays on Davidson's philosophy of action and psychology, including three by Davidson; excellent bibliography.)

Quine, W.V. (1959) 'Meaning and translation', in R.A. Brower (ed.) *On Translation*, Cambridge, MA: Harvard University Press, 148-71.(An earlier version of Chapter 2 of *Word and Object*.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(See Chapter 2, 'Translation and meaning', for discussion of radical translation.)

Quine, W.V. (1969a) 'Epistemology naturalized', in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 69-70.(Contains the argument for indeterminacy from holism and verificationism.)

Quine, W.V. (1969b) 'Ontological relativity', in *Ontological Relativity and Other Essays*, New York: Columbia University Press, 26-68.(Quine explains inscrutability of reference.)

Quine, W.V. (1970) 'On the reasons for indeterminacy of translation', *The Journal of Philosophy* 67 (6): 178-83. (Clarifies and expands upon earlier material.)

Quine, W.V. (1974) *Roots of Reference*, La Salle, IL: Open Court Press.(See pages 75-8 for discussion of barriers to translating truth functions.)

Quine, W.V. (1979) 'Comment on Newton-Smith', *Analysis* 39 (3): 66-7. (Quine expresses a reservation regarding the argument from underdetermination to indeterminacy.)

Quine, W.V. (1987) 'Indeterminacy of translation again', *The Journal of Philosophy* 84 (1): 5-10.(An excellent exposition of Quine's reasons for indeterminacy of translation.)

Quine, W.V. (1990) *Pursuit of Truth*, Cambridge, MA: Harvard University Press; revised edn, 1992.(Excellent

discussions of indeterminacy of translation and indeterminacy of reference.)

Quine, W.V. (1995) *From Stimulus to Science*, Cambridge, MA: Harvard University Press.(Excellent discussions of Quine's naturalism, reification and objects of thought.)

Wilson, N. (1959) 'Substances without substrata', *Review of Metaphysics* 12: 521-39.(Originator of the principle of charity.)

Reference

It is usual to think that referential relations hold between language and thoughts on one hand, and the world on the other. The most striking example of such a relation is the naming relation, which holds between the name 'Socrates' and the famous philosopher Socrates. Indeed, some philosophers in effect restrict the vague word 'reference' to the naming relation, or something similar. Others use 'reference' broadly (as it is used in this entry) to cover a range of semantically significant relations that hold between various sorts of terms and the world: between 'philosopher' and all philosophers, for example. Other words used for one or other of these relations include 'designation', 'denotation', 'signification', 'application' and 'satisfaction'.

Philosophers often are interested in reference because they take it to be the core of meaning. Thus, the fact that 'Socrates' refers to that famous philosopher is the core of the name's meaning and hence of its contribution to the meaning of any sentence - for example, 'Socrates is wise' - that contains the name. The name's referent contributes to the sentence's meaning by contributing to its truth-condition: 'Socrates is wise' is true if and only if the object referred to by 'Socrates' is wise.

The first question that arises about the reference of a term is: what does the term refer to? Sometimes the answer seems obvious - for example, 'Socrates' refers to the famous philosopher - although even the obvious answer has been denied on occasions. On other occasions, the answer is not obvious. Does 'wise' refer to the property wisdom, the set of wise things, or each and every wise thing? Clearly, answers to this should be influenced by one's ontology, or general view of what exists. Thus, a nominalist who thinks that properties do not really exist, and that talk of them is a mere manner of speaking, would not take 'wise' to refer to the property wisdom.

The central question about reference is: in virtue of what does a term have its reference? Answering this requires a theory that explains the term's relation to its referent. There has been a great surge of interest in theories of reference in this century.

What used to be the most popular theory about the reference of proper names arose from the views of Gottlob Frege and Bertrand Russell and became known as 'the description theory'. According to this theory, the meaning of a name is given by a definite description - an expression of the form 'the F' - that competent speakers associate with the name; thus, the meaning of 'Aristotle' might be given by 'the last great philosopher of antiquity'. So the answer to our central question would be that a name refers to a certain object because that object is picked out by the name's associated description.

Around 1970, several criticisms were made of the description theory by Saul Kripke and Keith Donnellan; in particular, they argued that a competent speaker usually does not have sufficient knowledge of the referent to associate a reference-determining description. Under their influence, many adopted 'the historical-causal theory' of names. According to this theory, a name refers to its bearer in virtue of standing in an appropriate causal relation to the bearer.

Description theories are popular also for words other than names. Similar responses were made to many of these theories in the 1970s. Thus, Kripke and Hilary Putnam rejected description theories of natural-kind terms like 'gold' and proposed historical-causal replacements.

Many other words (for example, adjectives, adverbs and verbs) seem to be referential. However we need not assume that all other words are. It seems preferable to see some words as syncategorematic, contributing structural elements rather than referents to the truth-conditions and meanings of sentences. Perhaps this is the right way to view words like 'not' and the quantifiers (like 'all', 'most' and 'few').

The referential roles of anaphoric (cross-referential) terms are intricate. These terms depend for their reference on other expressions in their verbal context. Sometimes they are what Peter Geach calls 'pronouns of laziness', going proxy for other expressions in the context; at other times they function like bound variables in logic. Geach's argument that every anaphoric term can be treated in one of these two ways was challenged by Gareth Evans.

Finally, there has been an interest in 'naturalizing' reference, explaining it in scientifically acceptable terms.

Attempted explanations have appealed to one or more of three causal relations between words and the world: historical, reliable and teleological.

1 Millian and description theories of proper names

The description theory of proper names stands in sharp contrast to the age-old and attractive theory that there is no more to a name's meaning than its role of designating something. Thus, John Stuart Mill claimed that 'proper names are not connotative: they denote the individuals who are called by them; but they do not indicate or imply any attributes as belonging to those individuals' ([1843] 1867: 20) (see Mill, J.S. §2). One problem for this theory is that it makes answering our central question seem hard: if a name does not 'imply any attributes' of its bearer, as the description theory claims it does, what determines which object is the bearer?

It was not primarily concern with this question, however, that led most philosophers to abandon the Millian theory and adopt the description theory. They did this largely as a result of the criticisms and counter-proposals of Frege (1893) and Russell (1911).

The most famous criticism of the Millian theory concerns identity statements. The ancient Greeks observed what they took to be a star rising in the evening and called it 'Hesperus', and what they took to be another star rising in the morning and called it 'Phosphorus'. In fact these 'two stars' were the planet Venus. Consider the following two statements:

(1) Hesperus is Phosphorus.
(2) Hesperus is Hesperus.

Sentence (1) is true. A comparison of (1) and (2) seems to indicate that they differ sharply in meaning. Various reasons have been adduced in favour of this view. Some philosophers have argued that whereas (1) is synthetic, (2) is analytic; others that whereas (1) is known empirically, (2) is known a priori; still others that whereas (1) is contingent, (2) is necessary. However, probably the most influential reason for thinking that the statements differ in meaning has been Frege's claim that they differ in 'cognitive value': (1) is highly informative, revealing an important astronomical discovery, whereas (2) is uninformative, a trivial piece of logical knowledge (see Frege, G. §3). In any case, it was generally agreed that the two statements do differ in meaning. If so, the only way to explain this seemed to be to attribute different meanings to 'Hesperus' and 'Phosphorus'. Yet, according to the Millian theory, they must have the same meaning as they both have the role of designating Venus. So the Millian theory must be wrong: a name's role of designating its bearer does not exhaust its meaning.

Another important criticism concerns existence statements. 'Vulcan does not exist' is true. Because it is true, 'Vulcan' does not designate anything. So, according to the Millian theory, 'Vulcan' should be meaningless. So, 'Vulcan does not exist' should be partly meaningless. Yet, it is perfectly meaningful. Indeed, if it were not, it could not be true. So, once again, the Millian theory must be wrong (see Fiction, semantics of).

The description theory provides neat solutions to what are problems for the Millian theory. According to the description theory, a name in effect abbreviates the definite description that competent speakers associate with the name: thus 'Aristotle' might abbreviate 'the last great philosopher of antiquity'. So it is easy to see why (1) and (2) differ in meaning: 'Hesperus' might be associated with the description 'the star that rises in the evening', and 'Phosphorus' with 'the star that rises in the morning'. And 'Vulcan does not exist' is fully meaningful because 'Vulcan' abbreviates the meaningful description 'the planet in orbit between Mercury and the Sun' (see Proper names §§1, 5).

More importantly for our purposes, the description theory provides an answer to our central question. A name designates a certain object because that object is denoted by the definite description associated with the name; 'Aristotle' designates Aristotle because 'the last great philosopher of antiquity' denotes him. Of course, this answer raises another question (see §6): in virtue of what does the description denote that object? Still, progress has clearly been made because we had that problem anyway.

Some obvious problems with this 'classical' description theory - including the problem that speakers differ in the descriptions that they associate with a name - led some philosophers, notably John Searle (1958) and Peter Strawson (1959), to modify the theory. A name is not tied tightly to one description but loosely to many. It can

designate its bearer despite the failure of some in its 'cluster' of associated descriptions to denote that object: it designates whatever object most of the descriptions in the cluster denote.

2 Three arguments against description theories of proper names

Description theories dominated for half a century until challenged by three arguments around 1970: the unwanted necessity and rigidity arguments, both due largely to Kripke (1980), and the argument from ignorance and error, due to Kripke and to Donnellan (1972).

Unwanted necessities were one of the obvious problems for the classical description theory. If 'the last great philosopher of antiquity' is synonymous with 'Aristotle', then 'Aristotle is a philosopher' should be necessarily true (provided Aristotle exists). Yet it is not: Aristotle might have died young, long before his philosophical fulfilment. The cluster theory avoided this version of the problem. The description 'the last great philosopher of antiquity' is just one among many in the cluster that expresses the meaning of 'Aristotle'. Aristotle need not have any particular one of the many properties specified by the cluster. The cluster theory does require, however, that Aristotle have *most* of the properties specified by the cluster. Kripke points out how implausible this is. Aristotle might not have had any of the properties commonly associated with him: he might not have been a pupil of Plato, taught Alexander the Great, and so on. So 'Aristotle' cannot be synonymous with the cluster of associated descriptions (see Kripke, S.A. §2).

The rigidity argument deploys the notion of 'rigid designation'. This is explained as follows: for a term *a* to be a rigid designator is for it to designate the same object in every possible world (in which it designates at all), or, less picturesquely, for it to be such that '*a* is *F*' would truly characterize some non-actual situation if and only if the object that the term *actually* designates were *F* in that situation. Kripke argues that names are rigid designators whereas the descriptions alleged to be synonymous with them are not. So description theories are wrong. Consider:

(3) Aristotle was fond of dogs.
(4) The last great philosopher of antiquity was fond of dogs.

Suppose that Aristotle had indeed died young. Then Plato, not Aristotle, would have been the last great philosopher of antiquity. In those circumstances the truth of (4) would depend on whether Plato was fond of dogs. But the truth of (3) would still depend, just as it does depend in the actual world, on whether Aristotle was fond of dogs. The name 'Aristotle' designates Aristotle in a non-actual situation just as it does in the actual situation, whereas the description 'the last great philosopher of antiquity' designates whoever is the last great philosopher of antiquity in that situation, whether Aristotle or not. So the name is not synonymous with the description. Similarly, it is not synonymous with any other description, or cluster of descriptions, that is a candidate to give its meaning. (Note that in assessing rigidity we are evaluating the truth and reference of our expressions with the meanings that they actually have as a result of our usage, but we consider them as used to characterize hypothetical situations. Of course, any expression could have a different meaning as a result of different usage in a non-actual situation - language is 'arbitrary' - but that is beside the point) (see Proper names §§2-3).

Statements (3) and (4) are not modal statements (although we have been evaluating them as characterizations of non-actual situations). Other versions of the rigidity argument concern modal statements. For example, whereas 'Hesperus is necessarily Hesperus' is true, 'Hesperus is necessarily the star that rises in the evening' is not: had the solar system been differently arranged, Hesperus might not have been visible in the evening but it still would have been Hesperus. This sort of difference between descriptions and names in modal statements had been emphasized earlier by other philosophers, particularly Ruth Barcan Marcus (1961).

Some philosophers, notably Michael Dummett (1973), resisted Kripke's two arguments by focusing on modal statements. These philosophers exploited the well-known ambiguities of scope in these statements to undermine the apparent difference between names and descriptions. Whatever the truth of this matter, the apparent difference in nonmodal statements remains.

These two arguments challenge the description theory as a theory of the meaning of a name, a meaning that determines the name's reference. This is how the theory is naturally understood. However, as Kripke points out, the theory could be understood as simply a theory of reference: the reference of a name is fixed by a description, but the name is not synonymous with that description. This weaker theory is impervious to the two arguments. Of

course, the weak theory has a defect: because it is no longer a theory of meaning, it no longer solves the problems that troubled the Millian theory. Indeed, the relation between meaning and reference becomes a pressing issue on the basis of this theory.

There is another way of saving the description theory from the two arguments while avoiding this defect. Instead of weakening the original theory into a mere theory of reference, we revise it along the following lines: a name is synonymous with a 'rigidified' description. Our language already seems to have descriptions that contain 'rigidity operators'; for example, the italicized part of 'the person who, *in the actual world*, is the last great philosopher of antiquity' seems to make this description designate Aristotle in every possible world. If descriptions of this sort are indeed rigid, the revised theory claims that a name is synonymous with such a description. If such descriptions are not rigid, the revised theory can claim that the name itself supplies the rigidity operator and so would be synonymous with an ordinary nonrigid description governed by that operator.

All of these description theories - original, weak and revised - have the consequence that the users of a name associate with it a description that identifies its bearer. The third argument against description theories, the argument from ignorance and error, challenges this. So, if the argument is good, it counts against all description theories.

The argument shows that people who seem perfectly able to designate with a name are very often too ignorant to supply an identifying description. Thus some may fail with the name 'Cicero' because they associate with it only the description 'a famous Roman orator', which applies to many people. Others may fail because they associate 'the man who denounced Catiline' with 'Cicero' and are unable to supply an appropriate description for 'Catiline': the description that they associate with 'Catiline' is 'the man denounced by Cicero', which takes us in a circle and leaves both names without reference, according to the description theory.

The argument shows also that people often associate with a name a description that identifies something other than the name's bearer: they are simply wrong about the bearer. Thus some associate 'the inventor of the atomic bomb' with 'Einstein' and some associate 'the first person to realize that the world was round' with 'Columbus'. Almost everyone who has heard of Peano associates 'the discoverer of Peano's axioms' with 'Peano', but the axioms were actually discovered by Dedekind. Despite such errors, people succeed in designating Einstein, Columbus and Peano by their names.

The description theory can be improved by allowing people to 'borrow' their reference from others. So the description Martha associates with 'Einstein' might be 'the person Joe referred to yesterday as "Einstein"'. Provided Joe can supply an appropriate description - either one that describes Einstein directly or one that borrows reference from someone who can supply an appropriate description - Martha will succeed in designating Einstein. There is a danger of a circle, of course. Apart from that, there are problems of ignorance and error once more. Perhaps Martha cannot remember the reference lender; or she can remember the lender by his name, 'Joe', but cannot supply the identifying description that the theory requires; or the lender is identified but he cannot identify Einstein, perhaps identifying something else instead. The description theory still seems to place too great an epistemic burden on speakers.

An argument from ignorance and error can also be brought against another, more general, theory that some - for example, Dummett - have taken from Frege. This is the theory that to understand a name a person must be able to 'identify' its bearer. This ability is usually evidenced by providing a description, but it may be evidenced by 'recognizing' the bearer. The epistemic burden that this more general theory places on speakers still seems too great.

Various moves have been made to save the description theory in the face of these difficulties. Most popular, perhaps, have been theories that the reference of, for example, 'Einstein' is determined by a description along the lines of 'the person referred to by (called, named and so on) "Einstein"', for this description does identify Einstein and speakers surely associate it with the name. However, such theories still risk circularity.

3 General terms and mass terms

Just as there are description theories of names, so also there are description theories of general terms like 'tiger', 'hammer' and 'bachelor', and of mass terms like 'gold' and 'paper' (see Mass terms). Speakers of the language

associate various descriptions with a term. One of these descriptions, or most of a cluster of them, expresses the meaning of the term and determines what it applies to. If only one description does the job, the view is analogous to the classical description theory of names. If a cluster of descriptions does, the view is analogous to the cluster theory of names.

Kripke (1980) and Putnam (1975) argued that description theories are false of general and mass terms that apply to natural kinds. So they are false of 'tiger' and 'gold'. The arguments are like the three against description theories of names (see §2). First, the theories yield unwanted necessities. The description we associate with 'tiger' is along the lines of 'large carnivorous quadrupedal feline, tawny yellow in colour with blackish transverse stripes and white belly'. Yet it is not necessary that a tiger has four legs and is striped: a tiger might lose a leg, or in a different environment tigers might not be striped. Second, the term 'gold' is a rigid designator, applying to the same kind of stuff in every possible world. In contrast, an associated description like 'dense yellow metal' is nonrigid. Third, people who seem perfectly able to use a term are often too ignorant or misguided about the things to which it applies to supply an appropriate identifying description. Thus, some who use 'elm' and 'beech' cannot supply descriptions that distinguish elms from beeches; many who use 'gold' cannot distinguish gold from fool's gold; it was once common to associate 'fish' with 'whale'.

Putnam added a further argument built around the following fantasy. Imagine that somewhere in the galaxy there is a planet, Twin Earth. Twin Earth, as its name suggests, is very like Earth. In particular, each Earthling has a *doppelgänger* on Twin Earth who is a cell for cell duplicate of the Earthling. Twin Earth differs from Earth in one respect, however: the stuff that the Twin Earthians who appear to speak English call 'water', stuff that is superficially indistinguishable from what we call 'water', is not H_2O but a very different compound XYZ. So Oscar on Earth and Twin Oscar on Twin Earth are referring to different stuff by 'water'. Yet Oscar and Twin Oscar are *doppelgängers*, associating exactly the same descriptions with 'water' (which is more plausible if we place Oscar and Twin Oscar in 1750 before the chemical composition of water was known). So those associations are not sufficient to determine reference and the description theory is wrong. Indeed, nothing happening in the head is sufficient to determine reference. As Putnam put it, 'meanings just ain't in the head' (see Content: wide and narrow; Putnam, H. §3).

We have considered criticisms of description theories of proper names and natural-kind words. Do these criticisms extend to description theories of other words? Putnam took the arguments to apply to almost all words, including 'pencil' and 'paediatrician'. Tyler Burge (1979) took a similar line, arguing that the meanings (contents) and references of a wide range of a person's words and accompanying thoughts are not 'individualistic' in that they are not determined simply by that person's intrinsic states. To a large extent they are determined by the person's social context. Burge's examples include 'arthritis', 'sofa', 'brisket', 'clavichord' and 'contract' (see Methodological individualism).

The Twin Earth fantasy brings out an important feature of description theories in general: even if a description theory gives the right answer to our central question for some word, its answer is incomplete. Thus, consider a description theory of 'tiger'. According to the theory, the reference of 'tiger' is determined by the reference of such words as 'carnivorous' and 'striped'. Suppose, contrary to the arguments above, that this were so. We then need to explain the reference of those words to complete the explanation of the reference of 'tiger'. Description theories might be offered again. But then the explanation will still be incomplete. At some point we must offer a theory of reference that does not make the reference of one word parasitic on that of other words. We need an 'ultimate' explanation of reference that relates some words directly to the world, if there is to be any reference at all.

4 Historical-causal theories

Kripke and Donnellan followed their criticism of description theories of names with an alternative view. This became known as the 'causal' or 'historical' theory, although Kripke and Donnellan regarded their view as more of a 'picture' than a theory (see Proper names §4).

The basic idea of this theory is that a name designates whatever is causally linked to it in an appropriate way, a way that does not require speakers to associate an identifying description of the bearer with the name. Reference is initially fixed at a dubbing, either by perception or description of the referent. The name is then passed on from

person to person in communicative exchanges. People succeed in designating an object with a name because underlying their uses of the name are causal chains stretching back to the dubbing of the object with the name. People borrow their reference from people earlier in the chain but borrowers do not have to remember lenders; it is enough that borrowers are, as a matter of historical fact, appropriately linked to their lenders in communication. So people can designate Cicero despite their ignorance of him, or designate Einstein despite their errors about him.

Similarly, Kripke and Putnam proposed an historical-causal theory of natural-kind words. Reference is initially fixed at a dubbing, either by description or perception of samples of the kind. The reference is then to all those objects, or all that stuff, having an internal structure of the same sort as the samples: for example, in the case of gold having the atomic number 79. People at a dubbing lend their reference to others, who can then lend it to still others. People who are ignorant about the kind can use the word to refer to the kind's members because underlying the use of the word are causal chains stretching back to a dubbing.

In thus removing the epistemic burden on speakers, historical-causal theories are a radical departure from the Frege-Russell tradition. That tradition assumes that those who understand a name must know about its meaning and reference, so that if its reference is determined in a certain way, they must know that it is. The historical-causal theory must reject the assumption that speakers have this privileged 'Cartesian' access to semantic facts: the reference of a name is determined by causal chains that are likely to be beyond the ken of the ordinary speaker. Once again, we must conclude with Putnam that 'meanings just ain't in the head'. This very feature of the theory has led many to reject it and to work hard to preserve the description theory (or the more general 'identification theory' favoured by Dummett; see §3). From the traditional Cartesian perspective, the causal theory's failure to impose an epistemic burden rules it out as a candidate to explain reference.

The historical-causal theory nicely captures the rigidity of names: the reference of a name is determined by its *actual* causal relations, something that cannot change when we consider other possible worlds. Less pleasingly, by rejecting any descriptive element to the meaning of a name, the theory may seem to leave no alternative but to resurrect the Millian view, identifying a name's meaning with its role of designating its bearer. Many philosophers, influenced by the 'direct reference' approach to indexicals (see §5), have taken this route, despite the problems for the Millian view (see §1). To avoid this it seems that we must explain a name's meaning in terms of the particular sort of causal chain that determines the name's reference.

The theory faces problems arising from various confusions and mistakes that can play a role in forming the causal network underlying a name. And it must explain how the reference of a name can change even though the historical fact of the dubbing cannot change. In developing the theory to deal with these problems, Michael Devitt (1981) has emphasized that a name is typically 'grounded' in its bearer in many perceptual confrontations after the initial dubbing: it is 'multiply' grounded in its bearer.

The historical-causal theory is returned to in §8.

5 Indexicals

To answer our central question for indexicals - terms like 'I', 'now', 'here', 'you', 'she', 'that' and 'this table' - we need to consider how the context of their utterance determines their reference.

Both Russell (1918) and Hans Reichenbach (1947) explain the reference of all indexicals (called 'egocentric particulars' by Russell and 'token-reflexive words' by Reichenbach) in terms of the reference of 'this'. Thus, Reichenbach claims that '"I" means the same as "the person who utters this token"; "now" means the same as "the time at which this token is uttered"; "this table" means the same as "the table pointed to by a gesture accompanying this token"' (1947: 284). This amounts to a description theory of reference for all indexicals except 'this token' in terms of the reference of 'this token'.

Any description theory of an indexical may face arguments of the usual three sorts: unwanted necessities, rigidity, and ignorance and error. The theory is particularly vulnerable to the argument from rigidity, as David Kaplan (1989) showed. (Kaplan prefers to talk of the closely related notion of 'direct reference' rather than 'rigidity'.) Thus, compare:

(5) This table is green

and Reichenbach's interpretation of it:

(6) The table pointed to by a gesture accompanying this token is green.

Suppose that the table referred to is in fact green, so that (5) and (6) both assert true propositions. Consider now a situation in which that table was still green but the furniture had been moved around so that a different table, a brown one, would be the subject of the gesture. Would those propositions asserted by (5) and (6) still be true? (Note that this question concerns the propositions actually asserted by (5) and (6), not the propositions that would have been asserted by the sentences in that non-actual situation.) Kaplan argues that whereas what (6) asserts would be false, what (5) asserts would still be true. The indexical 'this table' is rigid, referring to the same table in each possible world, whereas the description 'the table pointed to by a gesture accompanying this token' is nonrigid, referring to whatever table fits that description in the possible world. So the demonstrative is not synonymous with the description.

Apart from this, a general theoretical consideration counts against a description theory of indexicals. We have noted the essential incompleteness of description theories (see §3): even if a description theory is right for some word, the theory's explanation must rest ultimately on the reference of some other words which must be explained nondescriptively. Indexicals seem to be the most plausible candidates for nondescriptive explanation, more so even than proper names or natural-kind words: indexicals seem to be the place where language stands in its most direct relationship to the world.

In seeking a nondescriptive theory, it helps to follow Kaplan in dividing indexicals into two groups: 'pure indexicals' like 'I', 'here' and 'now', and 'demonstratives' like 'she', 'that' and 'this table'. The nondescriptive explanations of pure indexicals are fairly simple: 'I' designates the speaker of the utterance, 'here' the place of the utterance, 'now' the time of the utterance, and so on. (These explanations may seem to be description theories once again, but they are crucially different. For example, the last explanation is not that 'now' designates the time of the utterance because it is synonymous with an associated description 'the time of this utterance', but rather, because it is governed by the rule that it designates that time.) Demonstratives are more difficult to explain.

There are three basic ideas for a nondescriptive explanation of demonstratives. According to the first, a demonstrative designates the object demonstrated by the speaker. One problem with this idea is that a demonstration is often so vague that it alone would not distinguish one object from many others in the environment. A more serious problem is that demonstratives are not always accompanied by a demonstration. Thus, where only one table is prominent in the environment, the speaker may use 'this table' without a demonstration. And reference is often to an object that is not around to be demonstrated: for example, 'That drunk at the party last night was offensive'.

According to the second idea for a nondescriptive explanation, a demonstrative designates the object that the speaker intends to refer to. Even if this is so, it does not take us far because it raises the question: in virtue of what does the speaker intend to refer to that object? This is very similar to the original problem.

According to the third idea - urged, for example, by Edmund Husserl (1900-1) - a demonstrative designates the object in which it is based perceptually (compare the perceptually based grounding of a name according to the historical-causal theory). So 'this table' designates a certain table in virtue of the fact that it was perception of that table that led to the utterance; similarly 'that drunk at the party' designates the person that caused the remark (see Demonstratives and indexicals).

6 Descriptions

Definite descriptions have the form, 'the *F*', and indefinite descriptions the form, 'a/an *F*'. In his theory of descriptions, Russell (1905) claimed that 'the *F* is *G*' is equivalent to 'there is something that is alone in being an *F* and it is *G*'; and 'an *F* is *G*' is equivalent to 'there is something that is an *F* and it is *G*'. So the descriptions are to be understood in terms of the general term, '*F*' and the existential quantifier, 'there is something'.

Under the influence particularly of Donnellan (1966), many now think that a description is 'ambiguous', having not only this 'attributive' meaning captured by Russell but also a 'referential' meaning like that of a name or demonstrative.

It has been generally agreed that descriptions have a referential use as well as an attributive use. Used attributively, 'the F' conveys a thought about whatever is alone in being F; 'an F' conveys a thought about some F or other. Used referentially, each description conveys a thought about a particular F that the speaker has in mind, a thought about a certain F. Thus, consider:

(7) The murderer of Smith is insane

used in the following two contexts. (a) We come upon Smith foully murdered. We have no idea who is responsible but the brutal manner of the killing leads us to utter (7). Its description is used attributively. (b) We observe Jones on trial for Smith's murder. The oddity of his behaviour leads us to utter (7). Its description is used referentially. Next, consider:

(8) A man in a red cap stole Anne's computer

used in the following two contexts. (c) Anne's computer is discovered missing in the morning. We find signs that the burglar made a hasty escape dropping a red cap in the alley. This leads us to utter (8). Its description is used attributively. (d) After discovering that Anne's computer is missing we remember noticing a man in a red cap behaving suspiciously earlier in the day. We utter (8) to report our suspicions to the boss. Its description is used referentially.

Despite agreement that there are these two uses, there is no agreement that descriptions are ambiguous. Appealing to ideas prominent in the work of H.P. Grice (1989), many have defended Russell. They argue that a speaker can use a description referentially, thus making the object in mind the 'speaker referent', even though that object is not the semantic referent. Whether a speaker has an object in mind or not, the truth-conditions of the sentence are as specified by Russell. The referential use is pragmatically, but not semantically, different from the attributive use (see Pragmatics §§3-4, 8). Thus, in context (b), although Jones is the speaker referent, the truth of (7) will depend on the sanity of whoever murdered Smith, whether Jones or not. And in context (d), although the man behaving suspiciously is the speaker referent, the truth of (8) will depend on whether some man or other in a red cap stole the computer.

Against this, many have found reasons for thinking that in contexts like (b) and (d) the speaker referent is also the semantic referent and hence that descriptions are semantically ambiguous after all. Some of these reasons - for example, those arising from failures to describe correctly the object in mind and from the behaviour of descriptions in opaque contexts - have not stood up well. Others seem more promising. First, we are not just able to use descriptions referentially, it seems that we regularly do so. This regularity suggests that there is a convention of so using descriptions. If there is, then it is hard to see why the convention is not semantic. Second, in their referential uses descriptions seem to have roles just like demonstratives: 'the F' and 'an F' function like 'that F', and are similarly based on perception of a particular object. To try to treat these demonstratives like Russellian descriptions would be to give a description theory of them, and we have already noted problems for this (see §5). Third, definite (but not indefinite) descriptions seem to have the same range of anaphoric roles as a pronoun like 'she' (see §7). We might then expect them also to share the pronoun's role as a demonstrative which, as has already been pointed out, seems not to be Russellian. Fourth, consider the utterance, 'The book is on the table'. In the right circumstances, this will seem true and yet, according to the Russellian view, it must be false: since the world is full of books and tables the two definite descriptions fail to describe unique objects. The obvious modification to save the Russellian view is to treat these 'incomplete' descriptions as elliptical. But this modification has problems. A speaker may have many ways to complete the description and there may be no basis for saying that any one is the correct way. Alternatively, trying to complete the description may lead to the familiar problems of ignorance and error.

Another argument against the ambiguity thesis appeals to rigidity. If referential uses of descriptions were semantically significant then, it is claimed, they should be rigid like names and demonstratives. Yet they do not seem to be. Consider the use of 'Smith's murderer' in context (b) above. 'Smith's murderer is insane' does not seem to be true in a world where Smith is alive and well, even if Jones is insane. However, this argument has a problem: the referential use of the demonstrative 'that murderer' would equally seem to fail this rigidity test in these circumstances (hence suggesting a need to revise Kaplan's claims about demonstratives; see §5). Yet a demonstrative surely has a semantically significant referential use. If so, then a description may have one too,

despite not being rigid (see Descriptions).

7 Other terms

Many terms that we have not discussed - like adjectives, adverbs and verbs - are naturally taken to refer. It is certainly no easier to explain reference for these terms than for the terms discussed, but we may hope that doing so will not pose sharply different problems (see Adverbs; Predication).

We must also consider sentential operators like 'and' and 'not', and quantifiers like 'every pen', 'some stones', 'most dogs' and 'few bachelors'. Perhaps these should be seen as largely syncategorematic (see Logical constants; Quantifiers). If we are prepared to accept the existence of certain abstract entities, however, we can take these expressions as referential also. Thus we can take 'and' as denoting a 'truth function' *conjunction* which is such that the sentence '*p* and *q*' is true if and only if '*p*' is true and '*q*' is true. The quantifiers involve a 'determiner' and a general term, and can be taken as applying to sets. Thus 'most dogs' involves the determiner 'most' and the general term 'dogs' and can be taken as applying to any set that contains more than half the dogs; and the sentence 'most dogs bark' is true if and only if there is such a set and 'bark' applies to all its members.

In virtue of what do these expressions have these referents? The most promising answer for the sentential operators has two stages. We start by describing the 'conceptual role' of the operator in deductive, inductive and practical inferences. For a token to denote 'conjunction' it must have the appropriate conceptual role. But in virtue of what should we assign to a token with that role the denotation 'conjunction' rather than, say, 'disjunction'? Because, under that assignment, deductive inferences are truth-preserving, inductive inferences are reliable, and so on. A similar line is presumably part of the answer for the quantifiers: 'most dogs' applies to any set containing most dogs (rather than, say, to any set containing a few cats) partly because of the reference of 'dogs' and partly because of the conceptual role of the determiner 'most' and the reliability of inferences. A worrying feature of both these answers is that they seem to make widespread irrationality impossible (see Semantics, conceptual role).

Finally, we must consider anaphoric terms. Pronouns, and even definite descriptions, often depend for their reference on other expressions in their verbal context. Thus 'one' in 'John owns a car and Alice owns one too' is (using Geach's term) a 'pronoun of laziness', going proxy for the noun phrase 'a car' in the preceding conjunct. And consider 'he' in 'John is happiest when he is alone' and in 'Every man knows a woman that he admires'. In the former sentence 'he' is naturally seen as coreferential with 'John', in the latter as 'bound by' the quantifier 'every man' and so functioning like a bound variable in logic.

Geach (1962) has argued that all anaphoric pronouns are either pronouns of laziness or bound by quantifier antecedents. Against this Evans (1985) has argued that some pronouns with quantifier antecedents are unbound. He calls these 'E-type'. (a) Consider:

(9) Few congressmen admire the president, and they are very junior.

If 'they' were bound by 'few congressmen', (9) should mean that few congressmen both admire the president and are very junior. But it does not. The first clause of (9) entails that few congressmen admire the president; the second that all of those who do admire him are very junior. (b) If 'they' were bound in (9), then we should be able to substitute any quantifier for 'few congressmen' and still make sense. But 'No congressmen admire the president, and they are very junior' does not make sense. (c) Last, consider:

(10) If many men come to the ball, Mary will dance with them.

The quantifier 'many men' could bind 'them' only if (10) meant 'Many men are such that if they come to the ball Mary will dance with them', which is not the natural reading. As a result of these considerations and others - particularly pronouns in one person's sentence that are anaphoric on quantifiers in another person's - it is generally agreed that Evans has identified a distinct type of pronoun.

However, Evans' view of this type has been challenged. He thinks that the reference of such a pronoun is determined in a Russellian way by a definite description that can be derived from its quantified antecedent; thus, the reference of 'they' in (9) is determined by 'the few congressmen who admire the president', and that of 'them' in (10) by 'the many men who come to the ball'. Because these are definite descriptions, for (9) to be true *all* the Kennedy admirers must be junior, and for (10) to be true Mary must dance with *all* the men who come. This

consequence does not seem to generalize. 'Some congressmen admire the president, and they are very junior' seems to be compatible with some other congressmen admiring him and not being very junior. The problem is more acute in singular cases: 'Socrates owned a dog and it bit him' seems to be compatible with Socrates owning another dog which did not bite him. Finally, there are the formidably difficult 'donkey sentences': 'Every man that owns a donkey beats it' and 'If John owns a donkey, he beats it'. On one reading, these sentences concern not simply the unique donkey of each donkey owner but *all* the owner's donkeys (see Anaphora).

8 Naturalizing reference

From a naturalistic perspective, reference must ultimately be explained in scientifically acceptable terms. Attempted explanations have appealed to one or more of three causal relations between words and the world: historical, reliable and teleological.

Historical-causal theory of reference. Kripke, Donnellan and Putnam (see §4) did not claim to be naturalizing reference, but their theories - together with the role that perception of an object may play in determining the reference of demonstratives and referential descriptions (see §§5-6) - suggest the idea that reference might be explained naturalistically in historical-causal terms: a token refers to the object that played the appropriate role in causing it. But this idea, developed by Devitt, faces the *qua*-problem. In virtue of what is 'Aristotle', say, perceptually grounded in a 'whole object' and not a time-slice or undetached part of the object, each of which is equally present and causally efficacious? The problem is more pressing for natural-kind words. 'Horse' is grounded in a few horses. But those objects are not only horses, they are mammals, vertebrates and so on; they are members of very many natural kinds. Indeed, any horse is a member of indefinitely many *non*-natural kinds: it may be a pet, an investment and so on. In virtue of what is 'horse' grounded in an object *qua* horse, rather than *qua* mammal, pet, or whatever? So in virtue of what does it refer, as a result of such groundings, to all and only horses rather than all and only mammals, pets, or whatever?

Reliabilist theory of reference. Under the influence particularly of Fred Dretske (1981) and Jerry Fodor (1990), 'reliabilist', 'informational', or 'indicator' theories have been popular. The basic idea is that a token refers to objects of a certain sort because tokens of that type are reliably correlated with the presence of those objects; the tokens are 'caused by' those objects. The token 'carries the information' that a certain situation holds in much the same way that tree rings carry information about the age of a tree. There is a problem. How can the theory allow for error? Occasionally we see a muddy zebra and wrongly think 'horse'. So, some zebras are among the things that would cause tokens of 'horse'. What 'horse' is reliably correlated with is really the presence of horses, muddy zebras, the odd cow in bad light and so forth. So according to reliabilism, it should refer to horses, muddy zebras, the odd cow and so on (with the result that it was not wrong to think 'horse' after all). The problem is that many things that a token of a certain type does not refer to, including some denizens of Twin Earth, would cause a token of that type (see Semantics, informational).

Teleological theory of reference. Most fully developed by Ruth Millikan (1984), teleological theories explain the reference of a token in terms of its function, where that function is explained causally along Darwinian lines: a token's function is what tokens of that type do that explains why they exist. This theory deals neatly with the problem of error because something - for example, sperm - can have a function which it does not reliably perform. An immediate consequence of the theory is that a token of a type that has not evolved will lack a referent. So the 'thoughts' and 'utterances' of an exact replica of Russell created by some cosmic accident would have no reference. This strikes many as implausible but is accepted by the theory's proponents. To complete the theory it must be shown that tokens - even a belief like 'computers make writing easier' which could not plausibly be taken as innate - have a function in the required biological sense and that this function does indeed relate the token to its referent. Millikan has attempted this formidable task (see Semantics, teleological).

9 Further issues

Terms in opaque or intensional contexts cannot be seen as having their usual referential roles. For, in these contexts, particularly those of propositional attitude ascriptions (see Propositional attitude statements), the replacement of a term by a coreferential term may not preserve truth.

There are a range of what might be called 'negative' views of reference. (a) Some philosophers have a

'deflationary' view according to which there is nothing more to referential notions than is captured by all instances of a schema like '*e* designates *a*', where what is substituted for '*a*' 'translates' what is named by the term substituted for '*e*'; '"Socrates" designates Socrates' is a typical instance. This view accompanies a similarly deflationary view of truth (see Truth, deflationary theories of). (b) W.V. Quine (1960) argues that even once the translation of a sentence has been fixed the reference of any part of the sentence is inscrutable; thus there is no fact of the matter whether an alien's 'Gavagai' in response to an environment of rabbits refers to rabbits, undetached rabbit parts, time-slices of rabbits, and so on (see Radical translation and radical interpretation §§1-4). Related to this, Donald Davidson (1984) takes an instrumentalist attitude to reference, denying both the need for, and the possibility of, a theory of reference. Putnam (1983) gives a model-theoretic argument that reference is indeterminate because any theory has unintended models. (c) Kripke (1982) presents an argument (which he finds in Ludwig Wittgenstein's discussion of rule-following) that the meanings and references of terms are not determinate (see Meaning and rule-following). (d) Less sweepingly, Hartry Field (1973) has argued that in some cases there is no determinate matter of fact whether a term refers to one thing or another and we should see it as 'partially referring' to both: for example, 'mass' as used by Newtonians does not determinately refer to either proper mass or relativistic mass but partially refers to both. (e) Finally, those in the 'structuralist' tradition reject reference, and hence its role in meaning, altogether. They apparently think that the only possible theory of reference is one according to which a word resembles what it refers to. But this theory is refuted by the fact that language is arbitrary, by the fact that anything could be used to mean anything. Reference is thus left as simply 'God-given', which is unacceptable (see Structuralism in linguistics).

Finally, views on reference can bear on realist notions about the external world (see Realism and antirealism). Putnam, for example, draws antirealist conclusions from his model-theoretic argument. And consider the consequences of a holistic description theory for scientific terms. When, in time, we come to replace one scientific theory with another, it is natural to think that part of the reason we do so is that the theory does not accurately describe reality. Combine this thought with the holistic view that the reference of each term in the theory is determined by its associations with all other terms in the theory, and we get the consequence that all terms in the theory fail to refer. So, it was a mistake to believe in the entities apparently referred to by that theory. Worse, it is probably a mistake to believe in the entities of our present theory, for that theory will surely be replaced in time too. So we should not be scientific realists. Indeed, these considerations lead Thomas S. Kuhn (1970) and others to constructivism, a radically relativistic antirealism: rather than saying that the replaced theory does not describe reality, they say that it describes *its* reality, a reality that only exists relative to that theory (see Constructivism). Each theory has its own reality and no sense can be made of scientific entities existing 'absolutely'. This line of thought can be resisted by rejecting the holistic description theory of reference in favour of a localist theory, perhaps one explaining reference in terms of causal relations to reality (see Holism: mental and semantic).

See also: Semantics

MICHAEL DEVITT

References and further reading

Burge, T. (1979) 'Individualism and the Mental', in P.A. French, T.E. Uehling, Jr. and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 10, *Studies in the Philosophy of Mind*, Minneapolis, MN: University of Minnesota Press.(Argues that reference for many terms is determined by the social context; referred to in §3.)

Chastain, C. (1975) 'Reference and Context', in K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science*, vol. 7, *Language, Mind and Knowledge*, Minneapolis, MN: University of Minnesota Press.(A long paper on singular reference and anaphora. A source for the view that indefinite descriptions have two uses.)

Davidson, D. (1984) *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 215-25.(Argues against the need for, and possibility of, theories of reference; referred to in §9.)

Devitt, M. (1981) *Designation*, New York: Columbia University Press.(Develops a causal-historical theory of proper names, demonstratives and referential descriptions along naturalistic lines; referred to in §§4 and 8.)

Devitt, M. and Sterelny, K. (1998) *Language and Reality: An Introduction to the Philosophy of Language*, Oxford: Blackwell, 2nd edn.(Provides an introductory discussion of reference and of its significance for issues such as realism. Also includes an annotated guide to further reading.)

Donnellan, K.S. (1966) 'Reference and Definite Descriptions', *Philosophical Review* 75: 281-304.(Influential work proposing a distinction between referential and attributive descriptions and criticizing Russell and Strawson; referred to in §6.)

Donnellan, K.S. (1972) 'Proper Names and Identifying Descriptions', in D. Davidson and G. Harman (eds) *The Semantics of Natural Language*, Dordrecht: Reidel.(Criticizes the description theory of proper names and proposes a historical theory; referred to in §§2 and 4.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Most influential proposal of a reliabilist theory of reference; referred to in §8.)

Dummett, M.A.E. (1973) 'Appendix: Note on an Attempted Refutation of Frege', in *Frege: Philosophy of Language*, London: Duckworth.(A defence of a Fregean view of proper names from Kripke's criticisms; referred to in §2.)

Evans, G. (1982) *The Varieties of Reference*, ed. J. McDowell, Oxford: Clarendon Press.(Defends 'Russell's Principle' that to think about an object a person must know which object is in question.)

Evans, G. (1985) *Collected Papers*, Oxford: Clarendon Press.(Chapters 4, 5 and 8 argue for the view that some anaphoric pronouns are 'E-type' rather than bound; referred to in §7.)

Field, H. (1973) 'Theory Change and the Indeterminacy of Reference', *Journal of Philosophy* 70: 462-81.(Argues that the reference of scientific terms is sometimes indeterminate and that they should be seen as 'partially referring'; referred to in §9.)

Fodor, J.A. (1990) *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.(Chapters 3 and 4 propose a reliabilist theory of reference appealing to 'asymmetric counterfactuals'; referred to in §8.)

Frege, G. (1893) 'On Sense and Reference', in P. Geach and M. Black (eds) *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 1952.(Classic work raising the problem of identity statements for a Millian theory of proper names and proposing the theory that names have 'senses'; referred to in §1.)

Geach, P. (1962) *Reference and Generality*, Ithaca, NY: Cornell University Press.(Argues that all anaphoric pronouns are either like bound variables or are 'pronouns of laziness'; referred to in §7.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.(Proposes the distinction between speaker meaning and semantic meaning that has been influential in discussions of definite descriptions; referred to in §6.)

Husserl, E. (1900-1) *Logical Investigations*, trans. J.N. Findley, London: Routledge & Kegan Paul, 1970, part 1, §26; part 6, §§1-5.(Proposes a theory of demonstratives based on perception; referred to in §5.)

Kaplan, D. (1989) 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals', in J. Almog, J. Perry and H. Wettstein (eds) *Themes from Kaplan*, New York: Oxford University Press.(Influential work arguing that the reference of a demonstrative is 'direct' not via a Fregean sense; referred to in §5.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.(Very influential work criticizing the description theory of proper names and natural-kind terms and proposing a causal 'picture'; first presented in lectures in the late 1960s; referred to in §§2-4.)

Kripke, S.A. (1982) *Wittgenstein on Rules and Private Language: An Elementary Exposition*, Cambridge, MA: Harvard University Press.(Argues from a consideration of rule-following that the meanings and references of terms are not determinate; referred to in §9.)

Kuhn, T.S. (1970) *The Structure of Scientific Revolutions*, Chicago, IL: Chicago University Press, 2nd edn.(In this brilliant work on scientific change, a description theory of reference leads to antirealist constructivism; referred to in §9.)

Marcus, R.B. (1961) 'Modalities and Intensional Language', *Synthèse* 13: 303-22. (Argues that proper names are Millian and that true identity statements involving them are always necessary; referred to in §2.)

Millikan, R. (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.(The most detailed proposal of a teleological theory of reference; referred to in §8.)

Mill, J.S. (1843) *A System of Logic*, London: Longmans, 1867.(Book 1, Chapter 2 proposes his theory of both 'general and singular' names - referred to in §1.)

Neale, S. (1990) *Descriptions*, Cambridge, MA: MIT Press.(A thorough defence of the Russellian approach to definite descriptions. Includes an excellent bibliography.)

Papineau, D. (1987) *Reality and Representation*, Oxford: Blackwell.(Proposes a teleological theory of reference.)

Putnam, H. (1975) *Mind, Language and Reality: Philosophical Papers*, Cambridge: Cambridge University Press, vol. 2, 196-290.(Very influential argument against description theories and for causal theories of natural-kind and other terms; referred to in §§3 and 4.)

Putnam, H. (1983) *Realism and Reason: Philosophical Papers*, Cambridge: Cambridge University Press, vol. 3, 1-25.(Presents his model-theoretic argument that reference is indeterminate; referred to in §9.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(Section 12 argues for the inscrutability of reference in the context of arguing for the indeterminacy of translation; referred to in §9.)

Reichenbach, H. (1947) *Elements of Logic*, London: Macmillan.(Section 50 proposes a theory of 'token-reflexive' words; referred to in §5.)

Russell, B. (1905) 'On Denoting', *Mind* 14: 479-93.(A classic paper proposing his famous theory of descriptions; referred to in §6)

Russell, B. (1911) 'Knowledge by Acquaintance and Knowledge by Description', in *Mysticism and Logic*, London: Allen & Unwin, 1917.(A classic source for the description theory of proper names; referred to in §1.)

Russell, B. (1912) *The Problems of Philosophy*, London: Oxford University Press, 1959. (Chapter 5 gives a brief and accessible account of Russell's theory of proper names and descriptions.)

Russell, B. (1918) 'The Philosophy of Logical Atomism', in R.C. Marsh (ed.) *Logic and Knowledge*, London: Allen & Unwin, 1956.(Discusses the referential properties of many terms, including 'egocentric particulars'; referred to in §§1, 5 and 6.)

Searle, J. (1958) 'Proper Names', *Mind* 67: 166-73.(A very clear presentation of a cluster description theory of names; referred to in §1.)

Strawson, P.F. (1950) 'On Referring', *Mind* 59: 320-44.(A critique of Russell's theory of descriptions.)

Strawson, P.F. (1959) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen, 180-3, 190-4. (Proposes a cluster description theory of names; referred to in §1.)

Religious language

The main philosophical interest in religious language is in the understanding of what purport to be statements about God. Can they really be what they seem to be - claims to say something true about a divine reality? There are several reasons for denying this. The most prominent of these stems from the verifiability criterion of meaning, according to which an utterance can be a statement that is objectively true or false only if it is possible to verify or falsify it empirically. It is claimed that this is not possible for talk about God. However, the verifiability criterion itself has been severely criticized. Moreover, many religious beliefs do have implications that are, in principle, empirically testable, though not conclusively.

If one is moved to reject the idea that statements about God are what they seem to be, they can be taken as expressions of feelings and attitudes, and/or as guides to a life orientation. To be sure, religious utterances can have these functions even if they are also genuine statements of fact.

If one believes there to be genuine true-or-false statements about God, there are still problems as to how to understand them. We can focus on the construal of the predicates of such statements - for example, 'made the heavens and the earth' and 'commissioned Moses to lead the Israelites out of Egypt'. There is a serious problem here because of two basic features of the situation. First, the terms we apply to God got their meaning from their application to creatures, particularly human beings. Second, God is so radically different from us that it seems that these terms cannot have the same meaning in the two uses. One possibility here is that all these terms are used metaphorically when applied to God, which obviously often happens ('The Lord is my shepherd'). But are there some terms that can be literally true of God? This may be the case if some abstract aspect of the creaturely meaning of a term can be literally applied to God. For example, if one aspect of the meaning of 'makes' when applied to one of us is 'brings about some state of affairs by an act of will', the term 'makes' with that particular meaning might be truly applied to God.

1 The topic

The title of this entry is a misnomer. There is no language peculiar to religion ('Do you speak English, French or religious?'), nor is religion restricted to any particular type of language. A more accurate term for the topic would be 'religious uses of language'. Here we find an enormous diversity. Worshippers engage in praise, thanksgiving, petition, confession, instruction and exhortation. Sacred writings contain cosmological speculations, fictional narratives, historical records, predictions, commandments, theological pronouncements and legal codes. In devotional literature there are biographical reminiscences, theologizing, rules for the spiritual life and descriptions of religious experience. Philosophers of religion have concentrated on a very restricted portion of this plenitude, namely on statements about God or, more generally, about the objects of religious devotion and worship.

There is more than one reason for this selectivity. First, it is the belief aspect of religion with which philosophy is most concerned. This is partly an occupational bias. Philosophy is largely taken up with a critical examination of beliefs, assumptions and presuppositions in all areas. When the inquiry takes a linguistic turn, it is deflected to the linguistic formulations of beliefs. Since the most central statements in a religion are statements about God, they get most of the philosophical attention. Second, the belief system of a religion underlies everything else. Religious believers pray as they do, worship as they do, lead, or try to lead, their lives as they do, take up the attitudes they do, because of what they believe about God. If we are concerned to evaluate a religion, we are well advised to evaluate the belief system of that religion, since its pluses and minuses will have implications for the evaluative status of the whole.

Philosophers raise questions of various degrees of generality about religious statements. Some pose difficulties of understanding. The Christian doctrines of the Trinity (God is three persons in one substance) and the dual nature of Christ (the divine and human natures of one person), the Buddhist doctrine of *nirvāṇa* as the ultimate human fulfilment (is it pure nothingness, or does it have a positive aspect?), and the Hindu doctrine of Brahman (the absolute undifferentiated unity that constitutes all reality) are famous cases. But philosophers have also been concerned with more general questions as to how religious statements are to be understood. Traditionally this investigation has centred on the question of how predicates are to be understood in their application to God, predicates such as 'made the heavens and the earth' and 'knows the inmost secrets of our hearts'. For our grasp of

the subject, 'God', presumably comes from some predicates or other. Even if reference to God is at least partly on the basis of experience of God (Alston 1989: ch. 5), still it is also at least partly on the basis of what predicates we apply to him. Thus the question of how to understand predicates as applied to God is the most fundamental issue here.

It has generally been assumed that most apparent statements about God really do have that status, that they are used to make truth claims about a reality that is what it is independently of us, our beliefs, attitudes and conceptual schemes. But in the twentieth century that assumption has frequently been questioned. Hence, before tackling issues about the understanding of theological predicates, the reasons that have been given for denying the credentials of religious statements will be considered, along with the alternative construals of such utterances that have been proposed.

2 Reasons for denying that there are genuine statements about God

These reasons can be ranked under three headings: *metaphysical* - if treated as statements that are true or false, they are all false; *epistemological* - we have no effective way of determining their truth value; and *semantic* - because of the previous epistemological criticism, sentences predicating properties of God do not satisfy necessary conditions for having the kind of meaning (factual meaning) that would fit them for being used to make statements that are true or false.

The *metaphysical* claim is based on a naturalistic or materialistic metaphysics that takes reality to be confined to the 'natural' order, that is, to the physical universe in space and time (see Naturalized philosophy of science §1). Since that leaves no room for God, as usually conceived in religion, all statements that purport to refer to such a being are false. To be sure, this is quite compatible with taking what appear to be religious statements to have that status; they just all happen to be false. But if, while embracing naturalism, one is still motivated to hang on to something like traditional religion, the only option is to give religious utterances some nonstandard interpretation.

This line of thought can be no better than the naturalistic metaphysics on which it is based. Though materialists can claim some support from the developments of modern science for the thesis that everything in the spatiotemporal universe is purely physical in nature, it is not clear that they have any significant reason for denying that there are realities of a different order altogether, such as God is typically taken as being.

The *epistemological* reason is that we lack sufficient grounds for supposing that religious statements are true. Whether this is so is an extremely complicated issue that is treated elsewhere (see God, arguments for the existence of; Religion and epistemology; Religious experience). But even if it is so, that only implies that they must be accepted on faith. If one feels uncomfortable with that, it would be another motive for holding that apparent religious statements are not what they seem.

The *semantic* approach draws a stronger conclusion from an epistemological claim like the foregoing. The principle on which this reasoning is based is the verifiability criterion of meaningfulness. According to this principle, a sentence has factual meaning (the kind of meaning that renders it usable to make a statement with a truth value) only if it is in principle possible to verify it or falsify it empirically, on the basis of observations (see Meaning and verification §§2-3). The argument is that alleged statements about God fail this test and hence are not genuine statements of fact. In a famous passage, Antony Flew (1955) posed the rhetorical question, 'What would have to occur or to have occurred to constitute for you a disproof of the love of, or of the existence of, God?'. The implied answer is that nothing would fit this bill, and, by the same token, nothing would amount to a proof either. Alleged statements about God are only pretend-statements because they are not empirically testable.

It is this argument against the genuineness of statements about God that has received the most press. Again, it can be no stronger than the principle on which it is based, and the verifiability criterion has repeatedly been severely criticized (see Plantinga 1967). But in any case, do statements about God really fall foul of the requirement? That depends on a number of things. First, it depends on what counts as an 'observation' or 'empirical datum'. It makes a big difference whether mystical experience is allowed to count as 'observation'. Second, it depends on the viability of arguments for the existence of God, such as the teleological argument, that are based on observable features of the universe. And, on the other side, many thinkers take it that massive and apparently undeserved suffering constitutes empirical evidence against the existence of God. Third, it depends on details of the particular religious belief system in question. Many such systems involve fairly straightforward predictions as to what the

gods will do under certain conditions, where these divine actions manifest themselves in sensorily observable changes. In many primitive religions it has been believed that the gods will bring rain or military victory if they are approached through certain rites. In the Judaeo-Christian tradition it is believed that God will see to it that the Church or the chosen people will be finally victorious on earth and that prayers, made in the right spirit and in the right conditions, will be effective. It is true that such predictions come with severe qualifications. We cannot say *when* the Church will be victorious or when the second coming of Christ will inaugurate a new era. And it is impossible to be sure that one has prayed in the right spirit. Thus it is rare to find a decisive empirical test for religious beliefs. (But it is often held that this is true of science as well.) In any event, the issue of whether belief about God is empirically testable to some extent is by no means easily answered.

3 Nonassertive construals of theological statements

If one is convinced that talk of God can be meaningful only if it does not involve statements that are assessable as true or false, there are several options. Here are two:

Expressivism-instrumentalism. This is an analogue of 'noncognitivism' or 'emotivism' in ethics, a view that also springs from attachment to verificationism and which takes ethical utterances to be expressions of attitudes and emotions, or as recommendations of a policy of action, rather than as statements of fact. A classic version of the application to religion is found in Santayana (1905). On his view, there are two components to a religious doctrine, or 'myth'. There is (a) an evaluation of some sort, which is (b) expressed in the form of a picture or story. Thus the Christian myth of God's incarnation in Jesus Christ, and his death on the cross to atone for our sins can be regarded as an expression of the moral value of self-sacrifice. That is the expressive side of the position, but Santayana also thinks of religious myths as *guiding* our lives, our responses to the world. This side of the matter is indicated by the term 'instrumentalism', taken from the philosophy of science. The function for which religious beliefs are 'instrumental' is not predictive, as in science, but rather 'life-orienting' (see Braithwaite 1955).

Symbolisticism. Here a leading figure is Paul Tillich (§§3-5), according to whom it is misguided to ask whether religious doctrines are true or false of God. Since God (the true God, Being-Itself) is beyond any conceptualization (1953: 264-5), it is hopeless to seek any correspondence between what we say about God, and God. Instead, our 'God-talk' is made up of symbols of God, which 'point to' his reality by 'participating' in his power and being. Not only what are commonly recognized as symbols - the lamb, water, the shepherd, and so on - are to be so construed; anything concrete or conceptualizable is a symbol, including Christ, God the Father, and the Holy Spirit. Speaking of God the Father is an appropriate way of symbolizing Being-Itself because fatherhood is one of the 'places' in the world where we are 'grasped' by the power of Being. But any literal correspondence of our religious utterances with the divine is out of the question.

It must be noted that even if what seem to be factual religious statements do genuinely have that status, they can also function in an expressive-instrumental and a symbolic role.

4 Theological predicates

In discussing the question of how to understand predicates in their application to God we will assume that the arguments of §2 against the genuineness of statements about God are not cogent, and that such statements are to be understood as what they appear to be. Each such statement can be thought of as applying a certain predicate to God. Here is a sample:

(1) made the heavens and the earth;
(2) became incarnate in Jesus of Nazareth;
(3) is omnipotent (perfectly good, loving, wise, omniscient);
(4) forgives the sins of those that repent and turn to him;
(5) told me not to worry so much about trifles;
(6) wills that all people should be saved.

We can roughly divide this list into attributes (3), actions (1, 2, 4, 5) and intentional psychological states (6). The attributes can be seen as derivative from the other types, being properties of God that fit him for actions (loving, omnipotent) or for psychological states (omniscient, wise). Discussions of the topic have tended to concentrate on attributes, perhaps because they can be more concisely formulated, but this tends to distort the subject.

Why should there be a special problem here? We know what it is to tell someone something, to know something, to comfort someone, and so on. Is that not what we are saying of God in these cases?

There is a problem because, first, as the last paragraph suggests, we get our terms (concepts) for talking (thinking) about God from our talk and thought about creatures, particularly human beings, and, second, God is so different from human beings as to make the use of these terms problematic.

Regarding the first point, does it just happen that all the terms we use to specify God's attributes, actions and psychological states are terms we also use for human beings, or is there some deep reason for this? The latter alternative is supported by the following considerations. We have the kind of cognitive access to each other that makes it possible to establish a common language for talking about us. But we do not have the same resources for setting up an independent language for talking of God. To put it roughly, the parent can know when the child is perceiving another person talking or hugging someone or criticizing someone, and this makes it possible to introduce the child to the publicly shared meanings of those terms in application to human beings. But we cannot do anything analogous in the case of God. Even if the child can be aware of being spoken to or forgiven or comforted by God, the parent cannot tell when the child is aware of this unless the child tells the parent, which presupposes that the child already has learned how to apply the terms to God.

The second point is that creaturely terms cannot, in general, be used of God in exactly the same sense because of the ways in which God is different from creatures. There is no universal agreement as to just what these ways are, but some of those most commonly cited are as follows. God is infinite in power, knowledge and goodness; each of us is very limited in these respects. God is purely spiritual; we are embodied. God is omnipresent; we are severely restricted in spatial location. According to some theologies, God enjoys an atemporal mode of being - the simultaneous and complete possession of illimitable life; we live our lives successively, one moment at a time.

These differences have implications for what it is for God to be or do the various things we believe of him. God's knowledge is a very different thing from human knowledge. It is not built up by inference from the deliverances of perception. God has no sense organs, and no need to infer some truths from others. He knows everything directly. Since God is immaterial, divine overt action is a fundamentally different thing from human overt action. The only things we can effect directly are changes in our bodies; we bring about changes in the external world only by moving our bodies in certain ways. But God, having no body, will directly bring about changes in the world. Finally, if God's mode of being is atemporal, the picture will be still more different. Virtually every aspect of our activity is deeply imbued with temporality. We acquire information, think, deliberate and act by a temporal succession of stages.

This makes it clear that predicates cannot be truly applied to God in just the sense in which they apply to creatures. They cannot be used univocally of God and creatures. But that leaves several alternatives. They can be used metaphorically of God. That is very common in religious discourse. When we say 'His *hands* prepared the dry land' or 'The Lord is my *rock* and my *fortress*', the italicized terms are obviously used metaphorically. We do not think that God literally has hands or is a rock. And some people take all, or virtually all, theology to be metaphorical (see McFague 1982). But most theologians have supposed that we can make some literally true statements about God, and that it is important for the foundations of faith that we are able to do so. This has led them to explore other alternatives. A popular tack is that classically taken by Aquinas and developed in various ways by his successors. On this view, (some) creaturely terms are used analogically of God. There is an important analogy between their creaturely and theological import. According to Aquinas (§9), there is not even a partial univocity between 'know', for example, as applied to humans and God, even though there is enough analogy between their senses to enable us to understand what is said of God when he is said to know something. Aquinas' denial of partial univocity stems from his belief in a particularly radical divine-human difference, namely that God is absolutely simple. There is no difference of any sort between 'aspects' of the divine being. Since God is not different from any of his attributes or actions, no term as applied to God can mean even partly the same as to it does when applied to creatures (see Simplicity, divine §§1-2).

Contrary to Aquinas, there is a way in which, for some terms, there can be a partial overlap of meaning in the divine and human applications. The crucial point here is that the meaning of a term often includes more abstract, generic aspects and more concrete, specific ones. The term might carry over its more abstract meaning, though not

its more concrete meaning, from one application to another. This would make it possible for the more abstract component of the concept of knowledge to be common to God and humans, even though the way this is realized (and so the more concrete meaning of the term) would be different in the two cases. Consider an action term, such as 'makes'. It seems reasonable to suppose that the full concrete meaning of this term, as it is used of human beings in saying 'John made a bookcase', includes 'transforming some pre-existing material' and 'effecting this by moving his body in certain ways'. If so, that full meaning does not carry over to 'God made the universe', assuming that we mean this to involve creation *ex nihilo* and take God to be immaterial. But there could be a more abstract component of the meaning that is carried over, for example, 'brings it about by one or more acts of will that a certain state of affairs obtains'. In this way, some parts of the creaturely meanings of some terms can be used literally to make statements about God that have a chance of being true.

See also: Negative theology

WILLIAM P. ALSTON

References and further reading

Alston, W.P. (1989) *Divine Nature and Human Language: Essays in Philosophical Theology*, Ithaca, NY: Cornell University Press.(Philosophical essays on the nature of God and our talk of God.)

Aquinas, T. (1259-65) *Summa contra gentiles*, trans. V.J. Bourke, Garden City, NY: Doubleday, 1956, I, chaps 29-36.(See next item.)

Aquinas, T. (1266-73) *Summa theologiae*, London: Eyre & Spottiswoode, 60 vols, 1964, Ia, q.13.(This and the preceding item provide the classic statement of the view that terms applied to God have an analogical mode of meaning.)

Barbour, I.G. (1974) *Myths, Models, and Paradigms*, New York: Harper & Row. (Religious belief as based on models. Very readable.)

Braithwaite, R.B. (1955) *An Empiricist's View of the Nature of Religious Belief*, Cambridge: Cambridge University Press.(An interpretation of religious beliefs in terms of stories told for their moral point. Influenced by the verifiability criterion of meaningfulness.)

Crombie, I.M. (1957) 'The Possibility of Theological Statements', in B. Mitchell (ed.) *Faith and Logic*, London: Allen & Unwin.(A rich and complex view of the understanding of talk about God.)

Flew, A. (1955) 'Theology and Falsification', in A. Flew and A. MacIntyre (eds) *New Essays in Philosophical Theology*, London: SCM Press.(Classic application of the verifiability criterion to religious discourse.)

McFague, S. (1982) *Metaphorical Theology*, Philadelphia, PA: Fortress Press.(A stimulating presentation.)

Plantinga, A. (1967) *God and Other Minds*, Ithaca, NY: Cornell University Press.(A classic in philosophy of religion. Advanced.)

Santayana, G. (1905) *Reason in Religion*, New York: Charles Scribner's Sons. (An early account of religious discourse as non-factual.)

Tillich, P. (1953) *Systematic Theology*, vol. 1, London: Nisbet & Co.(Sets forth his view of theological discourse as symbolic; see also his 'Theology and Symbolism', in F.E. Johnson (ed.) *Religious Symbolism*, New York: Harper & Bros, and 'Religious Symbols and Our Knowledge of God', *Christian Scholar* 38 (3): 189-97, both published in 1955.)

Rhetoric

Rhetoric is the power to persuade, especially about political or public affairs. Sometimes philosophy has defined itself in opposition to rhetoric - Plato invented the term 'rhetoric' so that philosophy could define itself by contrast, and distinctions like that between persuasion and knowledge have been popular ever since. Sometimes philosophy has used rhetorical techniques or materials to advance its own projects. Some of its techniques, especially topics of invention, the classification of issues, and tropes or figures of speech, are occasionally employed by philosophers. The philosophical question is whether these techniques have any interest beyond efficacy. What is the relation between techniques effective in persuading others and methods for making up one's own mind? Is there any connection between the most persuasive case and the best decision? Is there a relation between the judgments of appropriateness and decorum exercised by the rhetorician, and the judgments of appropriateness exercised by the person of practical wisdom? Do judgments about probability, ambiguity and uncertainty, and judgments under constraints of time or the need for decision, aspire to the ideal of perfect rationality, to which they are doomed to fall short, or do these kinds of judgment have an integrity of their own?Apart from supplying useful techniques, an art of persuasion also raises philosophic questions concerning the relation between rhetoric and logic, rhetoric and ethics, and rhetoric and poetics.

1 Philosophy and rhetoric at their beginning

Plato invented the term 'rhetoric' as a contrast term against which philosophy could define itself. Philosophy and rhetoric both proposed new truths and apparently powerful methods that threatened existing moral codes and authorities. As modes of empowerment, self-making and self-consciousness they could as easily become enemies as allies. Thus while philosophy attempted to achieve self-consciousness and power through an awareness of thought and being, rhetoric focused on an awareness of language and the circumstances of speaking and acting. Cicero interpreted the Platonic separation of philosophy from rhetoric as that of Socrates dividing wisdom from eloquence, and saw that the task of philosophical rhetoric was to overcome this separation. Prior to Socrates, he says in *de Oratore*, 'the same teaching seems to have imparted education both in right conduct and in good speech…the same masters gave instruction in both ethics and rhetoric' (see Cicero §§1, 2). From the beginning, then, rhetoric already had the pejorative connotation it continues to have, in spite of periodic attempts to speak in its defence.

Rhetoric's importance for philosophy seems greatest in periods of political and cultural confrontation and pluralism, for example, during Roman expansion, the Renaissance and today. These are all times of 'linguistic turns' in philosophy and, in addition, of fruitful interactions between philosophic and legal argumentation and interpretation. Linguistic, legal and pragmatic turns are rhetorical phenomena that direct attention to effective reasoned communication. In such times, the revival of rhetoric is motivated by a sense that philosophy itself has become too professionalized and too remote from human concerns. Rhetoric is then invoked, as in Cicero, to bring philosophy back down to earth. That neither Rome nor the Renaissance is considered a golden age of philosophy is probably a reflection of their rhetorical character.

Whether or not the history of philosophy is a series of footnotes to Plato, the history of the philosophical issues of rhetoric certainly can be read that way. The *Gorgias* begins by showing the difficulty in defining rhetoric. Gorgias finds it hard not only to define rhetoric logically, but to offer a definition in the more serious sense of keeping it within its proper sphere. In Socrates' hands the universality of rhetoric becomes a sign of emptiness, not power, and its flexibility turns from resourcefulness into slavishness. The conversation between Socrates and Polus make this philosophy of rhetoric the first example of 'applied ethics', in which one searches for ethical constraints on an activity. The conversation between Socrates and Callicles originates the tradition of 'professional ethics' which relates the internal norms of a practice to the sort of person the practitioner is.

Other Platonic dialogues amplify these considerations and raise further questions. The same rhetoric that appears in the *Gorgias* as a power of domination appears in the *Protagoras* in a context of mutual agreement. The most obvious use of rhetoric - arguing both sides of a question - can be either to exercise power unilaterally or to provide the basis of community, just as competitive Homeric virtues either underlie or are transformed into the cooperative virtues of the polis. In the twentieth century, Kenneth Burke (1950) highlights this shift by changing

the central occupation of rhetoric from persuasion to 'identification'. The relation between the competitive and cooperative facets of rhetorical argument sets another philosophic problem for rhetoric: Under what conditions can the power to argue both sides of a question become the power to uncover truth through the free competition or marketplace of ideas? (In *On Liberty*, Mill develops his vision of truth emerging through controversy by reference to Cicero, 'the second greatest orator in antiquity'.) The introductory conversation in the *Protagoras* between Hippocrates and Socrates raises questions about the relation between learning from the sophists and being a sophist oneself. The dialogue then presents Socrates in conversation with the sophists Protagoras, Hippias and Prodicus. Protagoras limits the domain of rhetoric to political questions, justice and civic virtue, while Hippias would expand it to cover the natural sciences, and Prodicus would extend it to literary interpretation. Whether rhetoric is restricted to political questions, whether there is a 'rhetoric of science' or further uses of rhetoric outside politics all become standard questions for rhetoric from then onwards.

The *Phaedrus* turns to a third possible alliance which defines rhetoric in relation to poetics. In addition to rhetoric and logic (what does the rhetorician know?) and rhetoric and ethics (what are the uses and abuses of this power, and what sort of person does one become in practising rhetoric?) there is the relation between rhetoric and poetry, another activity that either employs charms not available to more rational approaches or operates at a further remove from reality. The *Phaedrus* raises further questions concerning the relation between written and spoken language, and so between the production and reception of discourse, a line of investigation that eventually leads from rhetoric to hermeneutics. Questions from the *Gorgias* concerning rhetoric and knowledge are formulated in the *Phaedrus* in a pointed manner as the relation between rhetoric and sincerity and deception, or between persuasion, *eros*, beauty, and seduction. Burke's investigations of the relations between serious and playful uses of language are a modern rediscovery of this set of questions, as are the recent revivals of rhetoric in de Man and Derrida.

The *Phaedrus* ends with Socrates asking whether there is a rhetoric for philosophy - that is, whether there are modes of communication that are especially suited for philosophy. But Socrates' questions about written versus spoken language, and long speeches versus questions and answers, at the same time pose the parallel problem of whether rhetoric has a philosophy, whether it has ethical or metaphysical commitments. The answer is parallel to that for the analogous question about mathematics: does the mathematical enterprise inherently have a 'philosophy'. There seems to be a natural affinity between mathematics and one type of philosophy, namely Platonism. On the other hand there are as many philosophies of mathematics as there are types of philosophy. Similarly, there are affinities of rhetoric with scepticism and relativism, and there are as many philosophies of rhetoric as there are kinds of philosophy (see Plato §§4, 5).

2 Philosophical problems posed by rhetoric

Rhetoric is persuasive communication, especially about practice and politics. Whether rhetoric is a subject for philosophy depends on whether persuasion raises issues of its own apart from those exhausted by logic, ethics and psychology. If the way people were persuaded had nothing at all to do with logical cogency or ethical responsibility, or if it was simply reducible to one of them, there would be nothing philosophic about rhetoric. It is only because there seems to be some connection between logical cogency and rhetorical persuasiveness, between the trust accorded a persuasive speaker and the confidence we have in good ethical agents, and between figurative language and something more than mere ornament, that rhetoric is a subject for philosophy. While rhetorical theory has had negligible influence on the history of philosophy, the problems raised by rhetorical practice are significant philosophical problems.

All the issues raised by Plato's dialogues could have become philosophical problems in a variety of ways, but they were disciplined by Aristotle's organization of knowledge. The details of Aristotle's *Rhetoric* have had little impact on the history of either philosophy or rhetoric, but his placement of rhetoric relative to other arts and sciences has organized the trajectory of the relation of philosophy and rhetoric. Rhetoric, he says, is the offshoot of dialectic and politics. There are three ends and kinds of rhetoric: political or deliberative rhetoric directed towards utility; judicial or forensic rhetoric which determines justice and injustice; and epideictic or demonstrative rhetoric which concerns worthiness and blameworthiness. There are three sources of persuasion or belief (*pistis*) in rhetoric - the character (*êthos*) of the speaker, the passions of the audience, and the 'speech (*logos*) itself, in so far as it proves or seems to prove' - generating the trio implicit in Plato: rhetoric and logic, rhetoric and ethics, rhetoric and

poetics.

The meaning and scope of all these terms of opposition (logic, ethics and poetics) shifts as the domain of rhetoric itself expands and contracts. Rhetoric is potentially about everything that can be communicated, and so about all thought and language. But the use of speech in deliberation about the concrete and practical particular is always at its core. As ethics after Aristotle became less political, so too did rhetoric. In Plato's dialogue, Protagoras defines his subject politically as the abilities without which men could not be citizens. Others softened the definition to the ability to speak on those matters on which people would be ashamed not to have an opinion, that is, the field of common sense and common understanding. Austin's (1962) inquiry captures part of the idea of persuasive communication apart from a political context, and shows the alliance between the linguistic and the rhetorical turns of philosophy (see Austin, J.L.). Some argue that rhetoric is as broad as human speech in general, with extensions to general rhetorics like the 'rhetoric of inquiry' or 'rhetoric of the human sciences' or, as with Burke, to all strategies for encompassing situations. Others would restrict it to public or political affairs. The broadening of rhetoric is sometimes a reflection of the lack of political, or even practical, functioning for rhetoric, and sometimes a corrective to the impractical specialization and professionalization of rhetoric itself. The logic, ethics and poetics against which rhetoric defines itself are usually much broader than the three sources of proof in Aristotle's account.

3 Rhetorical methods and philosophy

As rhetoric developed, it elaborated specific techniques for persuasion, and even if rhetorical activity had no philosophic dimensions, the history of rhetoric would still offer a usable history of persuasion. The first rhetorical method with philosophic import is that of topical invention, which brings into focus the relation between rhetoric and logic. *Topos* is the Greek word for place, translated into Latin as *locus*. *Koinoi topoi* became *loci communi* and then, in English, 'commonplaces,' places to find arguments, pigeon-holes for classifying appeals, or major premises from which to derive particular conclusions. Topics are the means for finding something persuasive to say in a practical situation. Aristotle defines rhetoric as an art of finding in any case the available means of persuasion. Subsequent rhetoricians make invention (*heuresis*, *inventio*) the first of five parts of rhetoric, alongside judgment (sometimes called disposition or arrangement), style, memory and delivery. Any of these can be a merely verbal technique or an art with philosophical significance. The topics can range from the logical ('similar effects have similar causes') to more substantive considerations ('follow the money'; *cherchez la femme*'; 'we all know that Athenians are acquisitive'). Often there is a connection between methods for finding arguments and systems of artificial memory for recalling them. When the rhetorical invention of arguments is replaced by a scientific discovery of things, only style remains in rhetoric's domain. Ramus' (1574) division of labour between a topical dialectic or logic and a rhetoric confined to style and arrangement in the early modern period is perhaps the most historically influential of such moves.

Scientific method might supplant rhetorical invention, but the methods of scientific discovery which replace rhetorical invention were themselves derived from topical invention. Much of the meaning of necessity in seventeenth-century science, for example, derives from pleas of necessity as an excuse or justification in legal proceedings. Stephen Toulmin's (1964) 'inference warrants', which supply the connection between grounds and claims, is one recent reappearance of the topics; Chaim Perelman's (1969) *loci* for establishing connections is another.

Hume finds the meaning of 'personal merit' through epideictic rhetoric by 'displaying the praises of any humane, beneficent man'. He refers to the considerations used as the 'topics of praise'. The inter-relations among the three kinds of rhetoric, and the kinds of argument appropriate to each, generate significant philosophical problems. Why, as Hume asks, does utility please? What is the relation between the impartial spectator whose judgment is the focus of epideictic rhetoric, the judge whose verdicts are the end of forensic rhetoric, and the ideal deliberative agent? What is the relation between the right and the good, or between what is best and what should be done? What is the connection between action-guiding and agent-evaluating reasons?

The other two philosophically interesting rhetorical methods besides topical invention can trace their ancestry back to Plato and Aristotle, but were really developed in later rhetorics which emphasized judicial oratory and the performance values associated primarily with demonstrative rhetoric. Issues - *staseis* in Greek and *constitutiones* in Latin - were rhetoric's central technique for most of its history. The classification of issues evolved out of

practices of judicial rhetoric, which is always the most systematic of the three kinds of rhetoric, and they draw attention to the relation between rhetoric and ethics. There are, usually, four *staseis*: (1) the question of fact or conjectural issue; (2) the definitive stasis; (3) the qualitative issue concerning mitigating or aggravating circumstances; and (4) the translative issue of whether a given court is the appropriate forum for judgment. Kant, not long after Hume, explains the 'principles of any transcendental deduction' by referring to stasis theory: 'Jurists, when speaking of rights and claims, distinguish in a legal action the question of right (*quid juris*) from the question of fact (*quid facti*)' (1787: B116-7). His question of right is the translative issue, by what right knowledge can be judged to have a 'legal title', so the foundations of knowledge cannot be secured by pointing to facts about the success of science. Austin's 'plea for excuses' and contemporary arguments about responsibility invite reconsideration of the rich problems that revolve around the qualitative issue.

If the topics emphasize the relations between rhetoric and logic, and the theory of issues or *staseis* develops the relation between rhetoric and ethics, the third rhetorical method, that of figures or tropes, concerns the relation of rhetoric to poetics. Augustine, in the *de Doctrina Christiana*, used rhetoric to teach Christians to read their Bible figuratively and to adjust contradictory passages by considering the circumstances to which speakers and authors adapted their statements and meanings. For Vico (1744), a near contemporary of Hume and Kant, the 'tropes are corollaries of poetic logic'. Borrowing from a formulation that seems to originate with Ramus, Vico names four primary tropes: metaphor, metonymy, synecdoche and irony. These tropes are not merely elegant variations on things that could be expressed more directly, but reveal original truths hidden by more sophisticated and professionalized philosophy. In the hands of our own contemporaries such as Kenneth Burke (1950) and Hayden White (1973), the figures move from methods of description or redescription into methods of seeing and encompassing situations, as rhetoric is expanded from a verbal to a universal art. Perelman, similarly, constructs a relation between argumentative forms and figures of speech and thought. Vico has become a hero to those who think that progress will come in philosophy from being anti-Cartesian, and championing conversation and communication over more limited, anti-rhetorical, forms of rationality. These examples show how even the most language-centred or poetic dimension of rhetoric has an ethical agenda. The rhetorical emphasis on saying what is appropriate provides the connection between ethical and poetic demands for timeliness and decorum. The current revival and popularity of rhetoric in philosophy often takes the form of a programme to replace science by conversation as the model for community and rationality and the hope for democracy.

See also: Language, medieval theories of; Language, Renaissance philosophy of; Legal reasoning and interpretation

EUGENE GARVER

References and further reading

Aristotle (*c.* mid 4th century BC) *On Rhetoric: A Theory of Civic Discourse*, trans. with introduction, notes and appendices by G.A. Kennedy, New York and Oxford: Oxford University Press, 1991.(The most readable and reliable English translation available.)

Augustine (*c.*396-427) *De Doctrina Christiana* (*On Christian Doctrine*), trans. D.W. Robertson, Jr, Indianapolis, IN: Bobbs-Merrill, 1958.(The most influential work explicitly adapting rhetoric to Christian religious and theological purposes.)

Austin, J.L. (1962) *How to do things with words*, Cambridge, MA: Harvard University Press.(A recovery of the rhetorical dimension of language under the title 'performative utterance'.)

Barthes, R. (1970) 'L'Ancienne Rhetorique: Aide-mémoire', *Communications* 16 (1970): 172-229; 'The Old Rhetoric: An Aide-Mémoire', in *The Semiotic Challenge*, trans. R. Howard, New York: Hill & Wang, 1988, 11-94.(Designed as a summary of the history of rhetoric for students of contemporary literary criticism.)

Burke, K. (1950) *A Rhetoric of Motives*, New York: Prentice Hall.(A cataloguing of the formal techniques for persuasion.)

Burke, K (1965) *Permanence and Change: An Anatomy of Purpose*, Indianapolis, IN: Bobbs-Merill.(An analysis of uses of poetry to solve ethical conflicts through rhetoric as 'an art of appeal'.)

Cicero (55 BC) *de Oratore*, trans. E.W. Sutton and H. Rackham, London: Loeb Classical Library, 1959.(Cicero's most philosophical treatment of rhetoric. Not a practical handbook but a dialogue on the place of rhetoric in ethical and political life.)

Conley, T.M. (1990) *Rhetoric in the European Tradition*, New York and London: Longman; Chicago, IL: University of Chicago Press, 1994.(A textbook, full of summaries and bibliographies; the best one-volume

history of rhetoric.)

Garver, E. (1994) *Aristotle's Rhetoric: An Art of Character*, Chicago, IL: University of Chicago Press.(A philosophical analysis of Aristotle's Rhetoric.)

Hume, D. (1748, 1751) *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press, 1966.(A moral theory based on the inference patterns of epideicic rhetoric.)

Kant, I. (1787) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan, 1929.(A critical philosophy rooted in justificatory rhetoric.)

Kennedy, G. (1980) *Classical rhetoric and its Christian and secular tradition from ancient to modern times*, Chapel Hill, NC: University of North Carolina Press.(One of a series of informative histories of rhetoric by this author.)

McKeon, R. (1987) *Rhetoric: Essays in Invention and Discovery*, Woodbridge, CT: Oxbow Press.(Contains the influential 'Rhetoric in the Middle Ages' and other articles by an important but difficult historian and philosopher of rhetoric.)

Mill, J.S. (1859) *On Liberty*, New York: Bobbs-Merill, 1956.(Chapter 2, in particular, concerns the power of controversy and discussion to reveal truth.)

Perelman, C. and Olbrecht-Tyteca, L. (1969) *The New Rhetoric: A Treatise on Argumentation*, Notre Dame, IN: University of Notre Dame Press.(A treatment of the argumentative force of figures of speech and the topics or *loci communes*.)

Plato (*c.*390-347BC) *The Collected Dialogues of Plato*, ed. E. Hamilton and H. Cairns, Princeton, NJ: Princeton University Press, 1971.(All Platos's dialogues collected in one volume, with an introduction and prefatory notes.)

Ramus, P. (1574) *Logikē*, trans. M. Roll, London: Scholar Press, 1966.(The only edition available in English of this highly influential work.)

Richards, I.A. (1936) *The Philosophy of Rhetoric*, New York and London: Oxford University Press.(Proposes a new rhetoric that studies misunderstandings and their remedies.)

Ricoeur, P. (1977) *The Rule of Metaphor*, Toronto, Ont.: University of Toronto Press. (A historical and philosophical investigation of the interrelations between rhetorical and poetic approaches to language and literature.)

Trimpi, W. (1983) *Muses of One Mind: The Literary Analysis of Experience and Its Continuity*, Princeton, NJ: Princeton University Press.(Traces the competition and joint contributions of philosophy and rhetoric to literary theory, from the Greeks to the Renaissance.)

Toulmin, S. (1964) *The Uses of Argument*, Cambridge: Cambridge University Press. (A re-discovery, under the name 'inference warrants', of the topics of classical rhetoric.)

Vico, G. (1744) *The New Science*, trans. T. Bergin and M. Fish, Ithaca, NY: Cornell University Press, 1984.(A defence of rhetoric against philosophy, and in particular the epistemic and ethical value of figurative over literal language.)

White, H. (1973) *Metahistory: The Historical Imagination in Nineteenth-Century Europe*, Baltimore, MD and London: Johns Hopkins University Press.(Uses of Kenneth Burke's theory of tropes to analyse historical narratives and philosophies of history.)

Sapir-Whorf hypothesis

The Sapir-Whorf Hypothesis is a widely used label for the linguistic relativity hypothesis, that is, the proposal that the particular language we speak shapes the way we think about the world. The label derives from the names of American anthropological linguists Edward Sapir and Benjamin Lee Whorf, who persuasively argued for this idea during the 1930s and 1940s - although they never actually characterized their ideas as an 'hypothesis'. In contrast to earlier European scholarship concerned with linguistic relativity, their approach was distinguished by first-hand experience with native American languages and rejection of claims for the superiority of European languages.

Early in the twentieth century, American anthropologist Franz Boas (1858-1942) inaugurated an important expansion of scientific investigation of the languages of native North America. As part of a broad critique of nineteenth-century evolutionary arguments he stressed the equal value of each language type and their independence from race and cultural level. He argued that each language necessarily represents an implicit classification of experience, that these classifications vary across languages, but that such variation probably has little effect on thought or culture.

His student Edward Sapir (1884-1939) accepted the main thrust of Boas' position but came to feel that the closely knit system of categories in a language could represent incommensurable analyses of experience with effects on speakers' conceptual view points and aesthetic interpretations. Gestalt and psychoanalytic psychology and Sapir's own literary efforts also played a role in his thinking on this issue. Sapir's concern was not with linguistic form as such (for example, whether a language uses inflections or not), nor with linguistic content or meaning as such (for example, whether a language could refer to a particular referent), but rather with the formal organization of meaning characteristic of a language, the regular ways meanings are constructed (for example, grammatical categories and patterns of semantic composition). Despite the suggestiveness of his formulation, Sapir provided few specific illustrations of the sorts of influences he had in mind.

Benjamin Lee Whorf (1897-1941), a gifted amateur linguist independently interested in these issues as they related to the nature of science, came into contact with Sapir in 1930 and began developing these views in a more systematic way. He analysed particular linguistic constructions, proposed mechanisms of influence, and provided empirical demonstrations of such influences on belief and behaviour. However, his views on this issue are known to us largely through letters, unpublished manuscripts and popular pieces, which has led to considerable debate about his actual position. In this context, the one article on this issue prepared for a professional audience must be given special weight (see Whorf 1956).

Whorf argued that each language refers to an infinite variety of experiences with a finite array of formal categories (both lexical and grammatical) by grouping experiences together as analogically 'the same' for the purposes of speech. These categories also interrelate in a coherent way, reinforcing and complementing one another, so as to constitute an overall interpretation of experience. Languages vary considerably not only in the basic distinctions they recognize, but also in the assemblage of these categories into a coherent system of reference. Thus the system of categories which each language provides to its speakers is not a common, universal system, but one peculiar to the individual language, and one which makes possible a particular 'fashion of speaking'.

But speakers tend to assume that the categories and distinctions of their language are natural, given by external reality. Further, speakers make the tacit error of assuming that elements of experience which are classed together on one or another criterion for the purposes of speech are similar in other respects as well. The crux of Whorf's argument is that these linguistic categories are used as guides in habitual thought. When speakers attempt to interpret an experience in terms of a category available in their language they automatically involve the other meanings implicit in that particular category (analogy) and in the overall configuration of categories in which it is embedded. And speakers regard these other meanings as being intrinsic to the original experience rather than a product of linguistic analogy. Thus, language does not so much blind speakers to some obvious reality, but rather it suggests associations which are not necessarily entailed by experience. Ultimately, these shaping forces affect not only everyday habitual thought but also more sophisticated philosophical and scientific activity. In the absence of another language (natural or artificial) with which to talk about experience, speakers will be unlikely to

recognize the conventional nature of their linguistically-based understandings.

The ideas of Sapir and Whorf have attracted widespread attention in the humanities and social sciences. Their views have been important in leading many students into comparative linguistics and played a crucial role in giving rise to the field of psycholinguistics in the 1950s. Despite this wide currency, their views have not been subjected to much empirical research. In large part, acceptance or rejection of their proposals has had more to do with the personal and professional outlook of the investigator and the prevailing temper of the times than with any solid evidence. In particular, philosophical responses, whether sympathetic or derisive, have rarely engaged with real linguistic phenomena (that is, how languages actually differ) or with practical cognition (that is, what kinds of effects one might expect). Instead debate centres on the logical (im)plausibility of linguistic incommensurability and determinism.

Existing empirical research on Whorf's claims has consisted primarily of tests for a relationship between lexical or grammatical categories (for example, words for colour, number marking) and experimental assessments of patterns of memory and classification. There have also been some attacks on Whorf's particular evidence, especially his claims about the Hopi language. This empirical research remains controversial, but some results clearly support Whorf's proposals and none decisively contradicts them. Another important strand of thinking has considered whether diverse uses of language (in particular the specialized discursive forms associated with language standardization, literacy and formal education) might have effects on thinking either in their own right or by mediating the structural effects proposed by Sapir and Whorf. This research often focuses on the cognitive importance of decontextualized speech and on the nature of speakers' conscious awareness and control of language - this latter an issue of concern to Whorf himself. Direct empirical research on such possible discursive effects remains quite limited.

There is also a growing body of research exploring the historical development of Sapir and Whorf's views, including intellectual biographies, more complete collections of their works, and accounts of the relation of their views to earlier European thinkers, especially in Germany. It seems certain that they were aware of this earlier work: Boas trained in Germany, Sapir wrote a Masters paper on Herder, and both Sapir and Whorf explicitly reject Humboldt's main substantive thesis, though not by name (see Herder, J.G.; Humboldt, W. von). However, their work is better understood, as with the rest of twentieth-century American anthropology, as an innovative departure from Europeanist views stimulated by direct first-hand contact with native American languages and cultures in a context of an emerging non-hierarchical view of culture.

See also: Austin, J.L.; Cassirer, E.; Condillac, E.B. de; Determinism and indeterminism; Diderot, D.; Hamann, J.G.; Language, philosophy of; Putnam, H.; Quine, W.V.; Radical translation and radical interpretation; Relativism; Searle, J.R.; Wittgenstein, L.

JOHN A. LUCY

List of works

Sapir, E. (1921) *Language: An Introduction to the Study of Speech*, New York: Harcourt, Brace & Co.(Classic treatment of language from an anthropological perspective; important for his early views; chapters 1, 4, 5, 10 and 11 deal with matters relevant to language and thought.)

Sapir, E. (1924) 'The Grammarian and His Language', repr. in D.G. Mandelbaum (ed.) *The Selected Writings of Edward Sapir in Language, Culture, and Personality*, Berkeley, CA: University of California Press, 150-9. (States his views on the diversity among languages as systems.)

Sapir, E. (1929) 'The Status of Linguistics as a Science', repr. in D.G. Mandelbaum (ed.) *The Selected Writings of Edward Sapir in Language, Culture, and Personality*, Berkeley, CA: University of California Press, 160-6. (States his view, especially on page 162, that language differences shape a speaker's view of reality.)

Sapir, E. (1931) 'Conceptual Categories in Primitive Languages', repr. in D.H. Hymes (ed.) *Language in Culture and Society: A Reader in Linguistics and Anthropology*, New York: Harper & Row, 1964, 128.(Dense, one-page summary of the shaping power of language at several levels.)

Sapir, E. (1933) 'Language', repr. in D.G. Mandelbaum (ed.) *The Selected Writings of Edward Sapir in Language, Culture, and Personality*, Berkeley, CA: University of California Press, 7-32.(Important for the explicit dismissal of nineteenth-century views of the Humboldtian sort regarding the correlation of language type, such as inflectional or isolating, with cultural type or level.)

Whorf, B.L. (1956) *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*, ed. and intro. J.

Carroll, Cambridge, MA: MIT Press.(The basic source for Whorf's writings on language and thought. Reprints 'The relation of habitual thought and behavior to language' his most important work for a professional audience (pages 134-59), several articles written for a popular audience (pages 207-70), and important unpublished manuscripts (pages 57-125). Carroll's introduction provides a brief intellectual biography and a bibliography of published and unpublished works.)

References and further reading

Aarsleff, H. (1988) 'Introduction', in W. von Humboldt, *On Language: The Diversity of Human Language-Structure and its Influence on the Mental Development of Mankind*, trans. P. Heath, Cambridge: Cambridge University Press, vii-lxv.(Useful summary of eighteenth- and nineteenth-century European views on linguistic relativity making clear en route their differences from contemporary views.)

Bertefflany, L. von (1955) 'An Essay on the Relativity of Categories', *Philosophy of Science* 22: 243-63. (Sympathetic treatment of the Sapir-Whorf hypothesis as analogous to other, commonly accepted types of incommensurability.)

Black, M. (1959) 'Linguistic Relativity: The Views of Benjamin Lee Whorf', *The Philosophical Review* 68: 228-38.(Influential article attacking Whorf's views as logically flawed.)

Boas, F. (1911) 'Introduction', *Handbook of American Indian Languages, Bureau of American Ethnology Bulletin* 40 (1): 1-83.(Classic statement of the Boasian view of language; see especially §2 on diversity among languages; various reprints available.)

Hymes, D. (1964) 'Two Types of Linguistic Relativity', in W. Bright (ed.) *Sociolinguistics, Proceedings of the UCLA Sociolinguistics Conference*, The Hague: Mouton, 114-65.(Early effort to consider the implications of work on language use or pragmatic function for Whorf's ideas.)

Gipper, H. (1972) *Gibt es ein sprachliches Relativitätsprinzip? Untersuchungen zur Sapir-Whorf-Hypothese*, Frankfurt am Main: S. Fischer; trans. R. Pinxton, 'Is There a Linguistic Relativity Principle?', in R. Pinxton (ed.) *Universalism versus Relativism in Language and Thought*, The Hague: Mouton, 1977, 217-28. (Contemporary German view on the Sapir-Whorf hypothesis, partly reproduced in translation.)

Gumperz, J.J. and Levinson, S.C. (eds) (1996) *Rethinking Linguistic Relativity*, Cambridge: Cambridge University Press.(Collection of contemporary anthropological research on issues related to renewed interest in the Sapir-Whorf hypothesis. Includes a review by Lucy of approaches suggesting relativity due to differences in language use.)

Koerner, E.F.K. (1992) 'The Sapir-Whorf Hypothesis: A Preliminary History and a Bibliographic Essay', *Journal of Linguistic Anthropology* 2: 173-8.(Useful brief introduction to the various historical studies of the Sapir-Whorf tradition and its various precursors. Good bibliography.)

Lee, P. (1996) *The Whorf Theory Complex: A Critical Reconstruction*, Amsterdam: John Benjamins.(Close intellectual history of Whorf and those who have worked on his ideas; makes heavy use of archival materials; extensive bibliography.)

Lucy, J. (1992) *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis*, Cambridge: Cambridge University Press.(Close textual analysis of the views of Boas, Sapir and Whorf and a critical analytic review of subsequent empirical research. Extensive bibliography of the social science literature.)

Lucy, J. (1992) *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis*, Cambridge: Cambridge University Press.(Outlines a method for investigating the Sapir-Whorf hypothesis and provides an extensive empirical investigation.)

Penn, J. (1972) *Linguistic Relativity versus Innate Ideas: The Origins of the Sapir-Whorf Hypothesis in German Thought*, The Hague: Mouton.(Traces important historical connections with earlier work.)

Rollins, P. (1972) *Benjamin Lee Whorf: Transcendental Linguist*, unpublished Ph.D dissertation, Harvard University, Cambridge, MA.(Biographical information on Whorf's intellectual interests. Volume 3 reprints some of Whorf's unpublished materials. The main arguments are reworked in *Benjamin Lee Whorf: Lost Generation Theories of Mind, Language, and Religion*, Ann Arbor, MI: University Microfilms, 1980.)

Silverstein, M. (1979) 'Language Structure and Linguistic Ideology', in P. Clyne, W. Hanks and C. Hofbauer (eds) *The Elements: A Parasession on Linguistic Units and Levels*, Chicago, IL: Chicago Linguistic Society, 193-247.(Influential extension of Whorf's arguments to questions of language use and native-speaker ideology.)

Scope

Scope is a notion used by logicians and linguists in describing artificial and natural languages. It is best introduced in terms of the languages of formal logic. Consider a particular occurrence of an operator in a sentence - say, that of ' →' in (1) below, or that of the universal quantifier ' ∀' in (2) below.

(1) A → (B & C)
(2) ∀x(Bxy → ∃ yAxy)

Speaking intuitively, the scope of the operator is that part of the sentence which it governs. The scope of ' →' in (1) is the whole sentence; this renders the whole sentence a conditional. The scope of 'Čamp;', on the other hand, is just '(B & C)'. In (2), the scope of the quantifier ' ∀' is the whole sentence, which allows it to bind every occurrence of x. The scope of ' ∃' is only ' ∃yAxy'. Since 'Bxy' is outside its scope, the 'y' in 'Bxy' is left unbound.

Although the importance of scope is semantic, it is usually identified with a syntactic relation between an occurrence of an expression in a sentence and a part of that sentence. In sentence and quantifier logic, the scope of an occurrence of an operator in a sentence S is usually defined as the smallest sub-sentence of S containing the occurrence. This ensures that for any two (occurrences of) operators with overlapping scope, exactly one is in the other's scope. In '$A →(B \& C)$', for example, 'Čamp;' is in the scope of ' →', and not vice versa. When the scope of one operator includes that of a second, the first is said to have wider scope than the second; the second's scope is narrower than that of the first.

Many semantically significant properties are characterized in terms of scope. The main logical operator of a sentence is the operator with the whole sentence as its scope. Thus conjunction, universal quantification, necessitation and so on are defined in terms of scope. A variable is bound whenever it occurs in the scope of a quantifier on it. A *de re* sentence is a sentence, such as ' $∃x$ I believe that x is a spy', in which a variable in the scope of a non-extensional operator (like 'believes' or 'necessarily') is bound by a quantifier outside that operator's scope. *De dicto* sentences are ones which do not involve such 'quantifying in' (see De re/de dicto).

Differences in relative scope can affect a sentence's truth-conditions. 'Many hate a few' has an interpretation, in which 'many' gets wide scope, on which it says that many have the property 'hating a few'. It has a quite different reading in which 'a few' gets wide scope, on which it says that a few have the property 'being hated by many'. If the quantifier 'someone' in 'I believe that someone is a spy' has a narrower scope than 'believes', the sentence does not imply that there is some individual whom I think is a spy; if the quantifier's scope is wider, it does imply this.

In a well-designed formal language, the relative scope of operators is indicated by surface syntax, but this is not inevitably true in natural language. A natural language sentence with n operators has potentially $n!$ (n factorial: $n(n-1)\ldots 3 \times 2 \times 1$) readings, since there are in principle that many ways of ordering the scopes of the operators. (Not all such orderings need be linguistically possible, of course.) Such ambiguities need to be resolved by an interpretive process, or a user's intentions, before the sentence can be evaluated for truth. Some syntactic theories posit a stage in sentence generation ('logical form') at which quantifier scope ambiguities are resolved by moving quantifiers so that their relative positions encode their scope (see Logical form).

A quantifier's scope should be the domain in which pronouns 'anaphoric on' (referring back to) the quantifier behave as variables bound by it. For example, the occurrence of 'her' in

(1) Every woman knew a man who loved her,

if anaphoric on 'every woman', behaves as a bound variable, since (1) is then understood as equivalent to

(1′) $∀x(x$ is a woman → x knew a man who loved x),

and the scope of 'every woman' is therefore the whole sentence.

Observe that only some uses of pronouns anaphoric on quantifiers behave like variables bound by those

quantifiers. For example, consider the sentence

(2) Garth bought some dogs and Wayne vaccinated them.

Even when anaphoric on 'some dogs', 'them' in (2) cannot be treated as a bound variable. If it were a bound variable, then the sentence as a whole would be equivalent to

(3) There are some dogs Garth bought which Wayne vaccinated,

but (3) does not imply that Wayne vaccinated *all* the relevant dogs, as (2) does. (Gareth Evans (1985) seems to be the first person to have noticed this.) So 'them' is to be excluded from the scope of the quantifier 'some dogs'. In accounts of natural language grammar which identify sentences with tree structures, quantifier scope is often identified with a structural relation between tree parts called c-command. While an expression in one sentence may be anaphoric on one in another, it is not possible for an expression in one sentence to c-command an expression in another. In particular, 'some dogs' in (2) does not c-command 'them'. It follows, on such accounts, that not all pronouns anaphoric on a quantifier are bound thereby.

Russell (1910-13) speaks of the scope of a description in giving his 'definition in use' for (that is, method for eliminating) definite descriptions. When the scope of a description 'the F in a sentence S is a sub-sentence 'T(the F)' of S, Russell says S is equivalent to the result of substituting (a formalization of) 'there is exactly one F, and T(it)' for 'T(the F)' in S. For example, if, in 'Possibly, the mayor is mad', the scope of the description is the whole sentence, the sentence is equivalent to 'There is exactly one mayor and possibly it is mad'. If the scope of the description is just 'The mayor is mad', the sentence is equivalent to 'Possibly: there is exactly one mayor and it is mad'. Russell's account makes the definite article semantically analogous to a binary quantifier.

Expressions other than quantifiers and connectives also have scope. Variable binders (such as function abstractors) and verbs which take sentential complements (for example, 'to believe') are examples. Standard definitions extend straightforwardly here.

One hears it said that proper names 'always take wide scope'. This seems intended as the claim that what is said by a use of a sentence $S(n)$, with n a proper name, is what would have been said by an appropriately related use of a sentence of the form 'n is such that S(it)', with 'it' anaphoric on n.

See also: Anaphora; Descriptions; Quantifiers, substitutional and objectual

<div align="right">MARK RICHARD</div>

References and further reading

Evans, G. (1985) *Collected Papers*, Oxford: Oxford University Press.(Essays 4, 5 and 8 discuss the interpretation of pronouns as bound variables. Somewhat difficult.)

Higginbotham, J. (1980) 'Pronouns and Bound Variables', *Linguistic Inquiry* 14: 679-708.(Proposes identifying Russell's notion of scope with c-command. Difficult.)

Neale, S. (1990) *Descriptions*, Cambridge, MA: MIT Press.(Discusses and defends Russell's account of descriptions. Also discusses contemporary accounts of binding and anaphoric relations. A good introduction to these topics.)

Russell, B. (1905) 'On Denoting', *Mind* 14: 479-93.(Russell's first presentation of his account of definite descriptions.)

Russell, B. and Whitehead, A.N. (1910-13) *Principia Mathematica*, Cambridge: Cambridge University Press, 3 vols; 2nd edn, 1925-7.(Volume 1 includes a formal treatment of descriptions and the scopes thereof.)

Semantics

Semantics is the systematic study of meaning. Current work in this field builds on the work of logicians and linguists as well as of philosophers. Philosophers are interested in foundational issues in semantics because these speak to the nature of meaning, as it embeds in our thinking and in our relations to each other and to the world. Of special interest are questions about how a semantic theory should respect the connections of meaning to truth and to understanding. In addition, numerous semantic problems concerning particular linguistic constructions bear philosophical interest, sometimes because the problems are important to resolving foundational semantical issues, sometimes because philosophical problems of independent interest are expressed using the constructions, and sometimes because clarity about the semantic function of the constructions enables clarity in the development of philosophical theories and analyses.

1 Historical sources and goals

Systematic theorizing about meaning in language always has been of philosophical interest. In addition to being a science with interesting foundational problems, semantics interests philosophers because of their special interests in the nature of meaning and in the meaning of particular sorts of language, notably the language that serves as our primary means of expressing philosophically interesting concepts and thoughts (see Language, philosophy of).

Contemporary semantics owes a great deal to work in formal logic by Gottlob Frege, Bertrand Russell and Alfred Tarski, among others (see Logic, philosophy of). This work has provided techniques for constructing *formal systems* - abstract systems of symbols and sets with rigorously defined logical and structural relationships (see Formal languages and systems). Ludwig Wittgenstein and logical positivists such as Rudolf Carnap found these techniques so compelling that they took certain formal, logical systems to reveal the kind of structure that any coherent language - in particular, any language suited to serious science - must possess at its roots (see Logical positivism). A less extreme but more enduring view is that many aspects of natural languages are aptly *modelled* by formal systems - whether formal systems devised independently by logicians or designed expressly to reveal the systematic structure of the languages in question.

Recently, semantics has profited from the development, under the aegis of Noam Chomsky, of systematic accounts of natural language syntax (see Syntax). Chomsky regards the grammar of a person's language as a partly innate theory that is implicitly mastered in acquiring the language (see Language, innateness of). The linguist's task is to make this theory explicit, guided by the person's spontaneous judgments about which constructions are grammatically acceptable, by experimental evidence about their cognitive faculties, and by general evidence about innate constraints on human grammars. Armed with this conception of the linguist's task, Chomsky and others have developed powerfully sophisticated theories of natural language grammar. Contemporary semanticists find this work useful for a number of reasons. Many semanticists now view their task in a Chomskian light: to uncover partly innate, cognitively real theories. Also, most contemporary semantical theories begin with something very like a grammar: a systematic account of the way sentences of the language in question are constructed from their meaningful parts - and it is quite reasonable to expect that the semantically relevant kind of sentence-structure bears a close relation to the correct grammar for the language.

2 Contemporary semantical traditions

The meaning of a statement is a matter of how things are, according to the statement; and the meaning of any expression is what users attach to it in their competent understanding of it. Semantics, then, ought to respect the outward-looking aspect of meaning: meaning fixes the representational and descriptive powers of language. And semantics must also respect the cognitive and epistemological aspect of meaning: meaning is cognized - it is grasped, known, understood.

A language (whether it be a shared social entity such as English, a narrower dialect, or even an individual's own 'idiolect') contains reusable words and expressions such as words and sentences. These are called expression-types because different tokenings of the expressions (different utterances or inscriptions of them) are importantly similar - they are of the same type (see Type/token distinction). Expression-types in a language have 'timeless' meaning that is drawn on in the various tokenings of the expressions. However, meanings can be expressed in the tokenings

that complete, or go beyond, the timeless meanings of the expressions used. For instance, a particular utterance of the sentence 'That was worse' would express a meaning to do with a particular subject matter ('that') and would involve (because of 'was worse') a gesture to another subject matter and to a dimension of evaluation. An attractive principle is that the timeless meaning of an expression simply is a matter of the expression's distinctive potential for expressing meanings in particular tokenings. The discipline of pragmatics (whose border with semantics is notoriously porous) seeks to understand the relations between the timeless meaning of expression-types and the meanings achieved by tokenings, as well as the variety of ways tokenings (or speech acts) can achieve communicative and other purposes (see Pragmatics; Presupposition; Propositions, sentences and statements; Speech acts).

A fact about natural languages that is no less impressive for being obvious is that every natural language issues in infinitely many sentences with different meanings. In English, 'Ann ate a pear' differs in meaning from 'Ann ate a pear, and then Ann ate a pear', which differs too from 'Ann ate pear, and then Ann ate a pear, and then Ann ate a pear', and so on. Any acceptable semantic account of a natural language will reveal its meanings as depending systematically on its repeatable, recombinant features (features of its expression-types and of their tokenings). One way to accomplish this is to show how meanings of complex expressions are generated from the meanings of simpler expressions, according to how the simpler expressions are combined. Such a strategy might employ a notion of *semantic values* of the meanings that are assigned to the simpler expressions (or tokenings) in the generation of the meanings of the more complex. That structure of a linguistic expression which, on these accounts, provides the key to the compositional derivation of its meaning sometimes is called the expression's *logical form*, because of assumptions about the nature of meaning to which we now turn (see Compositionality; Logical form).

The representational side of meaning has led many semanticists to take as their primary goal systematic accounts of the *truth-conditions* of sentences, as well as accounts of such related phenomena as the conditions of true application of predicates (such as 'is a horse') and the conditions of the denotation or reference of various noun phrases (such as proper names, definite descriptions like 'the last Tsar', and demonstratives like 'this'). It is immediately plausible that conditions of truth, applicability and reference are proper subjects for a semantic theory. For instance, the meaning of 'is a felucca', very plausibly, is a matter of what it is for a thing to be a felucca, which, again plausibly, is given by revealing what it takes for a thing to be truly described as a 'felucca'. Similarly, the meaning of 'In Syracuse there are many feluccas' is elucidated by characterizing how things would have to be to be truly so described (see Meaning and truth).

Because of the widespread view that a complete semantic account of a language must include systematic explanations of what it takes for its expressions to be true, to refer, or to apply truly, the central foundational issues in semantics include the nature of truth and reference, as well as the proper explanation and use within a semantic theory of the notions of truth- and applicability-conditions. Since truth- and application-conditions stand to each other in logical relations such as consistency and entailment, accounts of these conditions can serve what is regarded as another important goal in semantics: to explain the logical relations among linguistic expressions (and among their tokenings) (see Truth, coherence theory of; Truth, correspondence theory of; Truth, deflationary theories of; Truth, pragmatic theory of; Semantic paradoxes and theories of truth; Reference; Sense and reference).

A style of contemporary semantics that can be traced to Frege and Russell employs as semantic values abstract 'universals' such as *propositions* and *properties* - entities that embody truth- or application-conditions, in that (unlike linguistic expressions) they are not intrinsically tied to any particular language and have their truth- or application-conditions timelessly and essentially. Thus, some semantic frameworks start with conceptions of the nature of propositions and properties; within these frameworks the principal task in giving a semantical description of a given language is to discover the principles governing which expression-tokenings have as their semantic values which propositions and properties (see Propositions, sentences and statements; Universals; Intensional entities).

Some of these frameworks employ devices from the logical theory of models, in which the meanings of expressions are represented by set-theoretical structures (see Model theory). A particularly influential tradition in semantics deriving from the work of Carnap, Saul Kripke and Richard Montague, uses or imitates the notion of a 'possible world': the meaning of a statement is given by the sets of possibilities it excludes and allows, and the

meanings of smaller expressions are their contributions to this semantic function of the sentences in which they figure (see Semantics, possible worlds; Carnap, R.; Montague, R.M.; Kripke, S.A.; Semantics, situation). Among the philosophical insights claimed for this work is the idea that in addition to revealing a statement's simple truth-condition (the condition that the utterance must meet if it is to be true), a semantical theory ought further to reveal its *modal content*, which determines what sorts of possible situation the utterance correctly describes. Focus on this distinction has promoted intense debate about metaphysical necessity and possibility, in particular, on whether certain devices of natural language (such as proper names) presuppose substantive views about necessity and essence, and on whether these views are seriously tenable (see Proper names; Essentialism; Kripke, S.A.).

A somewhat different tradition in semantics has been pioneered by Donald Davidson, employing techniques due to Tarski (see Tarski's definition of truth). The goal in offering a Davidsonian semantics for a given language is to find an appropriate 'truth theory', thought of as a deductive system whose axioms and rules issue in 'T-sentences' such as 'The sentence "Schnee ist weiss" is true if and only if snow is white'. An appropriate truth-theory will not only be correct and complete, but also will in some sense explain the language-user's competence by revealing what it is that the speaker knows in knowing the language (see Meaning and truth; Meaning and understanding §2).

Recently, too, philosophers, linguists and cognitive scientists have developed systematic semantic accounts based on algorithmic or procedural conceptions of meaning. These traditions borrow from the logical theory of proof (see Semantics, game-theoretic; Proof theory).

All semantical traditions, even those that focus heavily on truth-conditions, must aim as well at least for compatibility with satisfactory explanations of the cognitive/epistemological aspects of meaning, including understanding (see Meaning and understanding; Meaning and verification; Sense and reference). One strategy is to keep issues about understanding at a remove from detailed semantic theorizing, by viewing understanding as involved not so much in the nature of meaning as in the uses to which meaning is put in thought and communication. David Lewis, for instance, proposes that a semantic account of a language is simply a correct description of the abstract function that assigns to statements in the language sets of possible worlds as meanings; understanding enters the picture only when we consider what constitutes the language's actually being used. For Lewis, the competent use of a particular language amounts to participation in a certain sort of conventional practice (see Language, conventionality of; Lewis, D.K.). Other strategies take semantic theorizing to be more directly focused on what the competent understanding of a language consists in. Some accounts in the Davidsonian tradition, for instance, submit truth-theories to constraints of cognitive plausibility: a truth-theory must not employ conceptual resources beyond those plausibly required of a competent language-user. Other accounts allow alien concepts in a semantic description of a language and employ a notion of *tacit knowledge* to explain how the semantic description represents what the language-user knows (see Knowledge, tacit). Also heavily cognitive in orientation are accounts within the Chomskian paradigm; these take the primary goal of semantics to be the discovery of the cognitive capacities and structures underlying linguistic competence (see Chomsky, N.). Some semantic accounts take as their task the explanation of how linguistic meaning is fixed by the meaning of mental items such as concepts and thoughts; for the ultimate source of meaning, these accounts defer to theories of cognitive semantics (see Mind, philosophy of; Grice, H.P.; Communication and intention; Concepts; Language of thought; Semantics, conceptual role; Semantics, informational; Semantics, teleological).

3 Issues and problems in semantics

Some debates about the semantic features of particular sorts of expression arise only within particular semantic traditions. Questions of *reference*, in contrast, are of very broad interest, since reference is widely taken to be a semantically central notion. Philosophers have hotly pursued questions about the nature of reference, about the proper role of reference in semantic theories and about how the reference of particular kinds of linguistic item is determined (see Reference; Sense and reference). Receiving special attention have been proper names, natural kind terms (like 'tiger' and 'gold'), context-dependent terms (such as the demonstrative 'that' and the indexical 'today') and theoretical terms of science (such as 'mass' and 'oxygen') (see Proper names; Natural kinds; Demonstratives and indexicals; Scientific realism and antirealism). Among the questions under heavy discussion are the extent to which the determinants of reference are settled by intrinsic features of a speaker's cognition, rather than being settled by features of the speaker's society and environment (see Language, social nature of; Content: wide and

narrow; Methodological individualism).

Issues about the semantical or logical structure of various locutions, in contrast, are often discussed in the context of particular traditions of semantic theory (see Logical form). Still, a number of such issues have interest that is not confined to any one semantic tradition. Among these are the proper treatment of definite descriptions such as 'the short spy'. Definite descriptions resist easy assimilation either to the paradigm of proper names or to that of quantifying phrases such as 'all circles' and 'three dogs' (see Descriptions; Quantifiers). Also of persisting interest is the behaviour of 'anaphoric' pronouns - ones which in one way or another 'refer back' (or even ahead) - such as occur in 'The dog is happy; he is wagging his tail' and 'Probably some dogs can climb trees; it is likely that they are small'. Difficulties in providing systematic explanations of the roles of anaphoric pronouns has led to a number of theories according to which semantic accounts of sentences must view their semantically relevant structure as inextricable from the larger discourse in which the sentences figure (see Discourse semantics; Anaphora). Even the logical form of apparently simple subject-predicate sentences has been debated. For instance, Davidson proposes that the logical form of 'Bill is running' involves not the simple predication of an attribute to Bill, but rather an existence claim about an event (perhaps: an event exists, which is a running, which is in progress, and of which Bill is the agent). This account is meant to cohere with a straightforward account of adverbs, in which, for instance, it is clear why 'Bill is running' is entailed by 'Bill is running slowly' (see Predication; Events; Adverbs).

So-called propositional-attitude ascribing statements (such as 'Sue believes that oranges are expensive') also have prompted extensive discussion. An ancient puzzle still has received no widely accepted solution: how is it that Hesperus and Phosphorus are one and the same entity, and yet 'Hammurabi believed that Hesperus is bright' might be true while 'Hammurabi believed that Phosphorus is bright' is false (see Propositional attitude statements; Indirect discourse; Intensionality)?

Another source of fascinating, but inconclusive, debate among semanticists is the treatment of conditionals (if/then sentences), including indicative conditionals like 'If Oswald didn't shoot Kennedy, then someone else did' and subjunctive, counterfactual conditionals like 'If Oswald hadn't shot Kennedy, then someone else would have'. Concerning indicative conditionals, one current view is that their truth-conditions are aptly modelled on those of the material conditional of formal logic, while another is that indicative conditionals have no truth-conditions at all but only conditions of proper assertion (see Indicative conditionals). Counterfactual conditionals have been linked to metaphysical issues about necessity, causation and natural law, and debates about their interpretation are enmeshed with metaphysical debates (see Counterfactual conditionals).

Also among the issues that bear on semantic theories are questions about necessity and possibility, the nature and ontological acceptability of intensional entities, tense, metaphor, vagueness and language apparently about nonexistent and fictional entities (see Modal operators; Modal logic, philosophical issues in; Intensional logics; Intensional entities; Tense and temporal logic; Metaphor; Vagueness; Existence; Fiction, semantics of; Fictional entities; Necessary truth and convention).

See also: Ambiguity; Analyticity; Emotive meaning; Imperative logic; Intuitionistic logic and antirealism; Logical constants; Mass terms; Meaning, Indian theories of; Meaning in Islamic philosophy; Ontological commitment; Questions; Semiotics; Structuralism in linguistics

MARK CRIMMINS

References and further reading

Larson, R. and Segal, G. (1995) *Knowledge of Meaning*, Cambridge, MA: MIT Press. (A lucid and philosophically sophisticated introduction to semantics developing a broadly Davidsonian approach.)
Chierchia, G. and McConnell-Ginet, S. (1995) *Meaning and Grammar*, Cambridge, MA: MIT Press.(An excellent introduction to semantics in the tradition of Montague grammar.)

Semantics, conceptual role

According to conceptual role semantics (CRS), the meaning of a representation is the role of that representation in the cognitive life of the agent, for example, in perception, thought and decision-making. It is an extension of the well-known 'use' theory of meaning, according to which the meaning of a word is its use in communication and, more generally, in social interaction. CRS supplements external use by including the role of a symbol inside a computer or a brain. The uses appealed to are not just actual, but also counterfactual: not only what effects a thought does have, but what effects it would have had if stimuli or other states had differed. Of course, so defined, the functional role of a thought includes all sorts of causes and effects that are non-semantic, for example, perhaps happy thoughts can bolster one's immunity, promoting good health. Conceptual roles are functional roles minus such non-semantic causes and effects.

The view has arisen separately in philosophy (where it is sometimes called 'inferential' or 'functional' role semantics) and in cognitive science (where it is sometimes called 'procedural semantics').

1 Motivations for CRS

There are two quite different projects that go by the name 'semantics'. One, which we might call *linguistic* semantics, deals with the meanings of particular expressions in particular languages and how they fit together to make up meanings of larger expressions. The second project, *metaphysical* semantics, is one of investigating the fundamental nature of meaning, especially what it is about a person that gives their words or thoughts whatever meanings they have in the first place. Conceptual role semantics (CRS) is in the domain of metaphysical semantics: it says that the nature of meaning is functional. It does not have anything very informative to say about linguistic issues, about particular languages or about how a language user works out the meanings of sentences on the basis of the meanings of their component words. But if correct, it can contribute to these enterprises by discouraging false and confused foundational views (see Semantics).

One major motivation for CRS is a functionalist approach to the mind generally (see Functionalism). Functionalism says that what makes a state a mental state and what gives a mental state the specific content that it has is the role it plays in interacting with other mental states in a creature's psychology.

This idea motivates a reply to theories that insist that a mind requires something more. For example, Searle (1980) has argued that computers cannot understand language in virtue of their programs or, more generally, by manipulating symbols in a certain way. He rests his case on a thought experiment, the Chinese room, in which a non-Chinese speaker manipulates Chinese symbols by following rules that do not require him to understand the meanings of the symbols he is manipulating. The rules are so devised that he produces sensible responses in Chinese to any Chinese inputs. Searle says that none the less he does not understand Chinese: he is just mindlessly manipulating symbols. CRS motivates the 'systems reply': if we can programme a computer to be intelligent, it will not be the central processing unit (CPU) all by itself that is intelligent or that understand the symbols, but rather all the complex relations between the CPU and other subsystems of the mind, for example, for perception, reasoning and decision making. So the whole system understands Chinese even if the person who is simulating the CPU does not (see Chinese room argument).

Approaching the matter from the point of view of language rather than thought, what makes CRS plausible is the fact that many terms seem definable only in conjunction with one another. For example, in learning the theoretical terms of Newtonian mechanics - 'force', 'mass', 'kinetic energy', 'momentum' and so on - we do not learn definitions outside the circle. There are no such definitions. We learn the terms by learning how to use them in our thought processes, especially in solving problems. Indeed, CRS explains the fact, noted by Thomas Kuhn (1962), that modern scientists cannot understand the phlogiston theory without learning elements of an old language that express the old concepts. The functional role of, for example, 'principle' as used by phlogiston theorists is very different from the functional role of any term or complex of terms of modern physics, and hence we must acquire some approximation of the eighteenth-century functional roles if we want to understand their ideas (see Definition; Scientific method).

Moreover, CRS does seem to give a plausible account of the meanings of the logical connectives. For example, we

could specify the meaning of 'and' by noting that certain inferences - for example, the inferences from 'p' and 'q' to 'p and q', and the inference from 'p and q' to 'p' - have a special status (they are 'primitively compelling', in the terminology of Peacocke 1992).

A further motivation for CRS is that it explains a reasonable version of a principle of charity according to which we cannot rationally attribute irrationality to a person without limit (see Charity, principle of §4). Attributing unexplainable irrationality leads to a poor match of roles. If the best translation yields poor enough matches, then the alien conceptual system is not intelligible in ours.

2 Two-factor CRS

Putnam (1975) raised what might seem to be a powerful objection to any CRS. He pointed out that many natural kind concepts, such as 'water' and 'gold', depend in part for their meaning upon something other than the role of a representation in a person's head, namely upon what happens to be in their external environment (see Content: wide and narrow; Methodological individualism).

Some proponents of CRS have responded by favouring a 'two-factor' version of CRS. On this view, meaning consists of an internal, 'narrow' aspect of meaning - which might be handled by functional roles that are within the body - and an external referential/truth-theoretic aspect of meaning, which might be handled by some other metaphysical theories of meaning (for example, a causal one). According to the external factor, 'Superman flies' and 'Clark Kent flies' are semantically the same since Superman = Clark Kent; it is the internal factor that distinguishes them. But the internal factor counts 'Water is more greenish than bluish' as semantically the same in my mouth as in the mouth of my twin on twin earth (see Content: wide and narrow §2); in this case, it is the external factor that distinguishes them.

Two-factor theories gain some independent plausibility from the need for them to account for *indexical* thought and assertions, assertions whose truth depends upon facts about when and where they were made and by whom (see Content, indexical). For example, suppose that you and I say 'I am ill'. One aspect of the meaning of 'I' is common to us, another aspect is different. What is the same is that our terms are both used according to the rule that they refer to the speaker; what is different is that the speakers are different. White (1982) generalized this distinction to apply to the internal and external factors for all referring expressions, not just indexicals.

In a two-factor account, the conceptual roles stop at the skin in sense and effector organs; they are 'short-arm' roles. But CRS can also be held in a one-factor version in which the conceptual roles reach out into the world - these roles are 'long-arm'. Harman (1987) has advocated a one-factor account which includes in the long-arm roles much of the machinery that a two-factor theorist includes in the referential factor, but without any commitment to a separable narrow aspect of meaning.

3 Criticisms of CRS

Error. Actual conceptual roles involve errors, even dispositions to err. For instance, in applying the word 'dog' to candidate dogs, one will make errors, for example, in mistaking coyotes for dogs (see Fodor 1987). This problem arises in one form or another for all naturalistic theories of truth and reference, but in the case of CRS it applies to erroneous inferences as well as to erroneous applications of words to things. Among all the conceptual connections of a symbol with other symbols, or (in the case of long-arm roles) with the world, which ones are correct and which ones are errors? Saul Kripke (1982), for example, wonders what distinguishes someone who mistakenly says '$57 + 65 = 5$' from someone who says it correctly, meaning by 'Čplus;' a function that agrees with addition except in yielding a value of 5 with 57 and 65 as arguments. The answer a person gives in the two cases could be the same, correct in one and erroneous in the other.

Some think we can solve the problem by appealing to dispositions to 'correct' previous answers, or to 'correct' those corrections. But others wonder why all these dispositions could not be the same for two persons who use 'Čplus;' to designate different functions. (The problem of error is sometimes said to be the problem of specifying semantic 'norms', although norms in this sense should not be confused with norms in the sense of how one ought to apply a word; see Horwich 1994.) Another line of reply is to attempt to specify some sort of naturalistic idealization which specifies roles that abstract away from error, in the way that laws of free fall abstract away from friction.

Words/world. Fodor criticizes a computer-oriented form of CRS for confusing what words denote with the words themselves. The functional roles in the target version of CRS stress searching data banks and manipulating representations, and this Fodor says is like claiming that the meaning of 'Napoleon won at Waterloo' is a set of instructions for finding that sentence in a book in the New York Public Library. All such a search yields is more words: we never get the semantic values of those words, namely Napoleon or Waterloo. But, the CRS theorist says in response, long-arm roles include causal chains outside the machine. And the two-factor version of CRS relies on a second factor, the referential factor, to explain the relation between the word 'Napoleon' and Napoleon.

CRS is often criticized from the point of view of truth-conditional theories of meaning (see Meaning and truth). If the meaning of a sentence is its truth-conditions, then the meaning cannot be its conceptual role. But with the two-factor theory, proponents of CRS have the option of counting meanings as the same or different in accordance with whether the external factor specifies truth-conditions that are the same or different. Further, there is reason to suppose that meaning is more fine-grained than truth-conditions. For example, the truth-conditions of 'I am happy' and 'Ned is happy' are the same (since I am Ned), but the meanings of those sentences differ. The further machinery involved in the internal factor can capture the differences among sentences with the same truth-conditions.

Sensory properties. Fodor also criticizes CRS for giving the wrong account of how I and Helen Keller (who was blind and deaf from an early age) can mean the same thing by, for example, 'Water tastes great'. After all, none of her thoughts bears the same relation to the evidence of sight and sound that mine do. But here Fodor assumes that CRS only has the resource of appealing to similarity in *inferential* role, which is entirely internal. He disparages such an account in favour of a referential view: we mean the same because our concepts of water are concepts of the same *thing*. But a two-factor CRS, relying in part on a referential component, has the option of giving exactly the same account as can a long-arm one-factor account.

What glues the two factors together. Fodor and Lepore (1992) object to the two-factor account, wondering what glues the two factors together. Why can there not be a sentence that has the inferential role of 'Water is greenish' but is true if and only if 3 is a prime number? But there is nothing in the CRS approach that dictates that there is any restriction at all on what roles can go with what truth-conditions. This is an independent question that both proponents and opponents of CRS can ask. Everyone who accepts the existence of inferential roles and truth-conditions should find the question meaningful, whether or not they think these are two factors of meaning.

4 Criticisms of CRS (cont.)

Holism. CRS is often viewed as essentially holistic, but the CRS theorist does have the option of regarding some proper subset of the functional roles in which an expression participates as the ones that constitute its meaning. Thus the subset could be taken to be those that are analytic (or 'true by virtue of meaning'); or as the primitively compelling inferences (Peacocke 1992) plus those generated by them; or the explanatorily basic regularities (Horwich 1994).

One natural and common view of what distinguishes the meaning-constitutive roles is that they are analytic, or played by an expression by virtue of its meaning, as in the case of an inference from 'bachelor' to 'male'. Proponents of CRS are thus viewed as having to choose between accepting holism and accepting that the distinction between the analytic and synthetic is scientifically respectable, a claim that has been seriously challenged by Quine (1954) (see Analyticity). Indeed, Fodor and Lepore (1992) argue that, lacking an analytic/synthetic distinction, CRS is committed to semantic holism, regarding the meaning of any expression as depending on its inferential relations to *every* other expression in the language (see Holism: mental and semantic). This, they argue, amounts to the denial of a psychologically viable account of meaning.

Proponents of CRS can counter as follows. First, there is a question of whether a meaning-constitutive inference is thereby analytic. *If* what is meaning-constitutive is analytic, then holistic versions of CRS need analyticity too, since they regard all inferences as meaning-constitutive. But if what is meaning-constitutive is *not* thereby analytic, then neither holistic nor non-holistic versions of CRS need analyticity. So analyticity is not the issue between holistic and non-holistic versions of CRS.

Second, proponents of CRS can reply that the view is not committed to regarding what is meaning-constitutive as

analytic. In terms of our earlier two-factor account, they can, for example, regard the meaning-constitutive roles as those that are explanatorily basic in a *narrow* psychology: they are the rules that explain other rules of use and determine *narrow* content (Horwich 1994). Narrow content does not involve *truth*-values; these arise only with regard to *wide* content, and so *a fortiori* it does not involve any commitment to *truth by virtue of meaning alone*.

A third approach to accommodating holism with a psychologically viable account of meaning is to substitute close enough similarity of meaning for strict identity of meaning. That may be all we need for making sense of psychological generalizations, interpersonal comparisons and the processes of reasoning and changing one's mind.

Compositionality. Fodor and Lepore (1992) raise a further worry that links the metaphysical semantic issue with a linguistic one: a CRS would seem to risk violating 'compositionality', that is, the requirement that the meaning of a complex expression be a function (in the mathematical sense) of the meanings of its parts (see Compositionality). It is widely thought that such a property of both language and thought is required to explain how human beings seem to able to grasp indefinitely many ever more complicated thoughts, and how they can learn to understand complex sentences on the basis of simple ones. CRS threatens this principle, since, Fodor and Lepore say, the conceptual role of a complex non-idiomatic representation is not always a function of the conceptual roles of its parts. Someone who thinks that rattling snakes, especially, are dangerous is disposed to infer 'This is dangerous' from 'This is a rattling snake' for reasons that may not depend at all on any inferences they are disposed to make from 'This is rattling' or 'This is a snake' separately.

Advocates of non-holistic versions of CRS should regard the argument's assumption that all inferences are to be included in inferential roles as question-begging. Non-holistic versions of CRS can deal with compositionality by counting only a subset of inferences as meaning-constitutive. As mentioned above, these inferences could be identified as the analytic ones, the explanatorily basic ones, or as those that are primitively compelling or generated by them. The threat to compositionality can be avoided by not counting the inference from 'This is a rattling snake' to 'This is dangerous' as part of the meaning-constitutive roles of either sentence.

Advocates of holistic versions of CRS may wish to go along with Fodor and Lepore in assuming that all inferences are part of inferential roles. They should point out that the inferential role of 'rattling' and 'snake' is a matter not just of their roles in *isolation* from one another, but also their roles in contexts involving 'rattling' and 'snake' *together*. The 'rules of use' of these terms are context-sensitive, not context-free.

Once we allow context-sensitive rules of use, compositionality can be trivially satisfied. For example, we can characterize the meaning of a word as an ordered pair, $\langle X, Y \rangle$, where X is the set of inferences *to* sentences containing the word and Y is the set of inferences *from* sentences containing the word. This is a holistic version of the view, for it includes the inference from 'rattling snake' to 'dangerous' in the meaning of 'rattling' and 'snake', and this example stands proxy for the inclusion of *every* inference in the meaning of every word involved in those inferences. Now the roles just mentioned satisfy the requirements of compositionality from a metaphysical point of view without being a psycholinguistic or a linguistic theory of the representations on the basis of which language is learned or sentences are understood.

5 Framework, not theory

CRS is more of a framework for a theory than an actual theory. There is no agreement among proponents of this framework about how the roles are constituted. By actual causal interactions among thoughts? All? Some? If some, which ones? And what about systematically mistaken inferences (for example, the 'gambler's fallacy')? Do widespread cognitive illusions contribute to the determination of meaning? Or are the roles normative? If the roles are idealized to avoid mistakes, how is the idealization supposed to be understood? Inference can be understood in intentional terms or in purely causal terms, and the latter would be preferable from the point of view of avoiding circularity in specifying roles. And is there any way to distinguish correcting an old practice from changing to a new one (Kripke 1982)? Many successful philosophical theories are quite sketchy. Some say that CRS is no worse than many of them, but others say that the problems in filling in these details involve difficulties that are fatal to the whole project.

See also: Concepts

NED BLOCK

References and further reading

Block, N. (1987) 'Functional Role and Truth Conditions', *Proceedings of the Aristotelian Society* 61: 157-81. (Argues that two-factor and one-factor theories do not differ substantively.)

Field, H. (1977) 'Logic, Meaning and Conceptual Role', *Journal of Philosophy* 69: 379-408.(Two-factor version of CRS based in conditional probability.)

Fodor, J. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press.(Raises the 'error' or 'disjunction' problem for theories of meaning.)

Fodor, J. and Lepore, E. (1992) *Holism: A Shoppers' Guide*, Oxford: Blackwell.(A critique of arguments for holism. Chapter 3 is concerned with CRS.)

Harman, G. (1987) '(Non-Solipsistic) Conceptual Role Semantics', in E. Lepore (ed.) *New Directions in Semantics*, London: Academic Press.(A defence of a one-factor CRS.)

Horwich, P. (1994) 'What it is Like to be a Deflationary Theory of Meaning', in E. Villanueva (ed.) *Philosophical Issues*, vol. 5, *Truth and Rationality*, Atascadero, CA: Ridgeview.(Shows how use theories can accommodate the truth-conditional, normative and compositional aspects of meaning.)

Kripke, S. (1982) *Wittgenstein: On Rules and Private Language*, Oxford: Blackwell.(Influential discussion of the 'error' problem, which Kripke claims to find in Wittgenstein.)

Kuhn, T. (1962) *The Structure of Scientific Revolutions*, Chicago, IL: University of Chicago Press.(A famous argument that the history of science is a series of routine periods punctuated by revolutions in which the scientific community changes its conception of the problems and of what the criteria are for a solution.)

Miller, G. and Johnson-Laird, P. (1976) *Language and Perception*, Cambridge, MA: MIT Press.(A CRS proposal in psychology.)

Peacocke, C. (1992) *A Theory of Concepts*, Cambridge, MA: MIT Press.(A CRS-oriented account of the nature of concepts.)

Putnam, H. (1975) 'The Meaning of "Meaning"', in K. Gunderson (ed.) *Minnesota Studies in the Philosophy of Science*, vol. 7, *Language, Mind and Knowledge*, Minneapolis, MN: University of Minnesota Press.(The source of the famous twin earth thought experiment.)

Quine, W.V. (1954) 'Carnap and Logical Truth', in *Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press; 2nd edn, 1976, 107-32.(Challenges the claim that the distinction between the analytic and synthetic is scientifically respectable.)

Searle, J. (1980) 'Minds, Brains and Programs', *Behavioral and Brain Sciences* 13 (4): 610-42.(Presentation of the 'Chinese room argument' against computational theories of mind and meaning.)

White, S. (1982) 'Partial Character and the Language of Thought', *Pacific Philosophical Quarterly* 63: 347-65. (The source of the view of narrow meaning based on Kaplan's theory of demonstratives.)

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Blackwell.(Source of the view that 'the meaning of a word is its use', an important inspiration for CRS.)

Woods, W. (1981) 'Procedural Semantics as a Theory of Meaning', in A. Joshi, B. Webber and I. Sag (eds) *Elements of Discourse Understanding*, Cambridge, MA: MIT Press.(A CRS proposal in artificial intelligence.)

Semantics, game-theoretic

Game-theoretic semantics (GTS) uses concepts from game theory to study how the truth and falsity of the sentences of a language depend upon the truth and falsity of the language's atomic sentences (or upon its sub-sentential expressions). Unlike the Tarskian method (which uses recursion clauses to determine satisfaction conditions for nonatomic sentences in terms of the satisfaction conditions of their component sentences, then defines truth in terms of satisfaction), GTS associates with each sentence its own semantic game played on sentences of the language. This game defines truth in terms of the existence of a winning strategy for one of the players involved. The structure of the game is determined by the sentence's structure, and thus the semantic properties of the sentence in question can be studied by attending to the properties of its game.

For each sentence *S* of the language, there is a semantic game G(*S*) between two players, the defender of *S* and the attacker. *S* is true when the defender has a way of winning the game, no matter what the attacker does - that is, when the defender has a winning strategy in G(*S*). *S* is false when the attacker has a winning strategy. The defending player defends a disjunction by selecting a disjunct to defend, since the truth of a disjunction requires only a single disjunct to be true. Since the truth of a conjunction requires both its conjuncts to be true, the attacker attacks a conjunction by selecting a conjunct for the defender to defend. This means that the defender must be able to defend both conjuncts, if the conjunction is to be true. To defend a negation $-S$, it suffices to show *S* false, so the defender of $-S$ is the attacker of *S*, and vice-versa.

A play of G(*S*) eventually comes down to a single atomic sentence *A*. Since the truth-values of nonatomic sentences depend on those of atomic ones, G(*S*) is played against the background of a specified assignment of truth-values to the language's atoms. If *A* is true according to the given assignment, the defender of *A* wins this play of G(*S*), but if it is false, the attacker wins. Truth of *A* is not necessary in order for the initial sentence *S* to be true; for instance, the defender may defend a disjunction unwisely by picking a false disjunct, even if the disjunction is true. *S* is true when the defender of *S* has a winning strategy, which means that there is a way to force any play of G(*S*) to end in victory for that player.

To show an existential quantification to be true, it suffices to find a single true instance. To show a universal quantification false, it suffices to find a single false instance. Thus the defender of an existential sentence selects an individual as an instance of the existential claim, and the attacker of a universal selects an instance of that universal. Again, the universal is true only if the defender can defend every instance that the attacker might select.

When applied to an ordinary first-order language (with no restrictions imposed on strategies available to the players), the assignment of truth-values to sentences produced by game-theoretic semantics (GTS) - on the basis of an assignment of values to the atoms - coincides with the results of the more familiar semantic frameworks. The usefulness of GTS appears when attention turns to natural language, and also when restrictions are imposed on available strategies in games for formal languages.

Unlike most formal languages, English is structurally ambiguous in ways that a semantics must treat (see Ambiguity). For instance, 'Everyone loves someone' is ambiguous between two readings, $(\forall x)(\exists y)(x \text{ loves } y)$ and $(\exists y)(\forall x)(x \text{ loves } y)$, with the former reading strongly preferred. This ambiguity arises in GTS because there are distinct games for the sentence. In one game, the universal quantifier 'everyone' is processed first, thus receiving a wide-scope interpretation so that the first reading is produced. In the second, the existential quantifier 'someone' is processed first. The first reading is preferred because 'everyone' commands 'someone' in the sentence. ('Command' refers to the structural relation of the two quantifiers: the first is 'higher' in the sentence than the second). Thus GTS for English involves a principle that commanding operators must be processed first. This is an 'ordering principle' which governs the structures of semantic games so as to produce preferred readings of ambiguous sentences. Some ordering principles can be absolute, others (like the example above) can be defeasible. Another defeasible ordering principle is that operators should be processed following the left-to-right order of their occurrence in the sentence in question. This principle also applies to 'Everyone loves someone'.

Unlike many approaches to the semantics of English, GTS is not compositional. The interpretation of a sentence need not be determined in a 'cumulative' way based on the predetermined meanings of its components. This

allows GTS to treat various semantic phenomena that do not respond to compositional semantic analysis (see Compositionality).

GTS applied to English yields interesting and novel treatments of such well-known semantic problems as anaphora (determining the conditions for co-reference of pronouns and related singular terms and quantifiers), conditionals, intentional identity, negation and nonstandard informational relationships among operators in a sentence. This can be seen in so-called 'branching quantifiers' and in the interpretation of relational questions. For example, the branching sentence:

$$\left.\begin{array}{l} \forall x \exists y \\ \forall z \exists w \end{array}\right\rangle Fxyzw$$

has *y* within the scope of *x* and *w* within the scope of *z*, but *y* is not within the scope of *z*, nor *w* within the scope of *x*. This pattern of quantifier relationships cannot be represented in a linear first-order quantifier prefix. The semantic game for this sentence is like the game for $\forall x \exists y \forall z \exists w Fxyzw$, except that the game for the branching sentence is a game of imperfect information. The defender of the sentence is ignorant of certain previous moves in the game, whereas the defender of the linear sentence knows what has transpired in the game for that sentence (see Quantifiers §2).

Branching quantification is a well-understood example of informational independence of operators. Informational independence is currently the object of intense study in GTS, which has led to the formulation of 'independence-friendly' logics and the discovery of conditions governing informational variability. Other areas currently of great interest include negation in English and the role of subgames in interpretation. The subgame of a semantic game is a self-contained game serving as a component of the larger game, so that the larger game involves results of the subgame in its play. For example, interpretation of a conditional requires information concerning the semantic game for its antecedent, in order to interpret pronominal connections between the antecedent and consequent. Recent developments have generalized the subgame idea, using 'tangent games' to provide a semantics of relative clauses and to link them to homophonic embedded questions.

See also: Decision and game theory; Semantics

MICHAEL HAND

References and further reading

Hand, M. (ed.) (1994) *Synthèse* 99 (3), special issue.(Special GTS issue containing important new work on negation, informational variability, games with chance moves and other topics.)

Hintikka, J. (1976) *The Semantics of Questions and the Questions of Semantics*, Amsterdam: Acta Philosophica Fennica.(An early, extensive semantic investigation of questions.)

Hintikka, J. (1983) *The Game of Language: Studies in game-theoretical semantics and its applications*, Netherlands: Kluwer.(Papers by Hintikka - some in collaboration with Kulas - on negation, tense and other topics.)

Hintikka, J. and Kulas, J. (1985) *Anaphora and Definite Descriptions*, Dordrecht: Reidel.(A GTS account of pronominal relations.)

Hintikka, J. and Sandu, G. (1990) *On the Methodology of Linguistics*, Dordrecht: Reidel.(Focuses largely on a GTS account of pronominal relations.)

Saarinen, E. (1979) *Game-Theoretical Semantics*, Dordrecht: Reidel.(A collection of early papers in GTS, including Hintikka's influential discussions of branching quantification, Hintikka and Carlson's original discussion of subgames in the semantics of conditionals, Saarinen on backward-looking operators and intentional identity, and Rantala's urn models.)

Semantics, informational

Information-theoretic semantics (ITS) attempts to provide a naturalistic account of the conditions under which a psychological state such as a belief or desire has a particular mental content: what it is by virtue of which, say, a psychological state is a belief 'that it is raining' or a desire 'that it stop raining'. Because of the complexities of an entirely general account, ITS typically attempts to provide merely a sufficient *naturalistic condition for a belief content of the sort normally acquired by perception (for example, that it is raining). It is expected that other sorts of mental contents may require that ITS be supplemented in various ways.*

ITS was inspired by Claude Shannon's theory of 'information' (1948), which provided a mathematical measure of the amount of information carried by a signal. Employing a notion of 'natural meaning' discussed by Peirce (1931) and Grice (1957), Dretske (1981) supplemented Shannon's work with an account of what *information a signal carries. The intuitive idea is that a signal carries the information 'that* p*' if and only if it naturally means (that is, indicates) that* p*, as when smoke 'means' there is fire.*

Natural indication is a key ingredient in ITS accounts of mental content. In their accounts, Stampe (1977) and Stalnaker (1984) appeal to the notion of what a state indicates under 'optimal' conditions. Fodor (1987) appeals to 'asymmetric dependencies' between the meaning-forming and the non-meaning-forming indication conditions in the causation of psychological states. Dretske (1988) appeals to the idea that, via operant conditioning, a state can acquire a functional role vis-à-vis *behaviour because it naturally indicates 'that* p*' and thereby can acquire the natural function of indicating 'that* p*'.*

1 Content and indication

In an early proposal, Dretske (1981) maintained that a state, *S*, carries the information 'that *p*' (that is, has the informational content 'that *p*') if and only if the probability of *S*, given *p*, is equal to 1 (see Information theory). One issue this proposal raises is whether the probability is 'objective' or 'subjective' (a matter of degrees of belief) (see Probability, interpretations of). If the notion of information content is to be suitable for a naturalistic account of mental content, the notion of probability invoked must be objective. But, as Dretske himself pointed out, the relevant notion of probability is not relative frequency. Moreover, it seems not to be propensity either: the direction of propensity is the direction of causation; but the conditional probability Dretske would have us consider is the opposite direction: the probability of there being 'fire', given that there is 'smoke'. In a later proposal, Dretske (1988) employed a counterfactual: the occurrence of a state, *S*, carries the information that *p* if and only if *S* would not have occurred unless *p*. He claimed that such counterfactuals express objective, mind-independent facts.

2 Two problems

Error. A state carries the information 'that *p*' if and only if it indicates (or naturally means) 'that *p*'. But a state indicates 'that *p*' only if *p* is indeed the case. So information, in the sense of natural indication, does not admit of misinformation, or error. However, belief does: one can believe 'that *p*' even when *p* is false. How can information-theoretic semantics (ITS) accommodate this fact?

Fine-grained individuation. Moreover, even when a belief 'that *p*' indicates 'that *p*', the indication relation will not uniquely pair the belief with *p*. The belief 'that *p*' can differ from the belief 'that *q*', even when *p* and *q* are logically equivalent. For example, the belief that 'there is a chair in the room' has a different content from the belief that 'there is a chair in the room and it is either an antique or not an antique', even though these contents are logically equivalent. However, if a state indicates *p*, then it indicates anything that is logically equivalent to *p*. Indeed, if a state indicates 'that *p*', then it indicates anything that is merely nomologically equivalent to *p*: for example, anything that indicates the presence of a renate (creature with kidneys) indicates the presence of a cordate (creature with a heart). Moreover, if something indicates 'that *p*', then it indicates anything that logically or even nomologically follows from *p*: anything that indicates that something is a cow indicates that it is a herbivore. Yet the belief that something is a renate is different from the belief that it is a cordate, a belief that something is a cow is different from the belief that it is a herbivore.

3 Three proposals

Optimal conditions. Stampe (1977), Fodor (1990b) and Stalnaker (1984) try to accommodate the possibility of error or misrepresentation by maintaining that the content of a psychological state is what the state would indicate under 'optimal' conditions. Stalnaker's proposal is that a belief state has the content 'that *p*' if and only if, in optimal conditions, the state indicates *p*, and the subject is in the state because *p* or because of something that entails *p*. The idea is that in cases of false belief, conditions are not optimal (see Idealizations).

One problem is that it is by no means clear that the notion of ideal conditions is suitably naturalistic. Belief fixation is holistic: what someone would come to believe in some situation depends on what they already believe about other things. Even in a good light, someone presented with a horse might not think it is one if they think appearances are misleading. If optimal conditions must include the subject's being in belief states with certain contents, then, as an account of belief content, ITS is circular.

Asymmetric dependencies. Fodor (1987) formulates his ITS theory as one for mental symbols - expressions in a 'language of thought' (see Language of thought). He calls tokens of mental symbols meaning 'cow' that are caused in the absence of cows 'wild', and calls the property whereby symbols can mean things that on occasion are not causes of their tokening 'robustness'. He points out that in solving problems of robustness and error, ITS needs to solve the 'disjunction' problem: given that among the causes of a symbol's tokenings, there are both meaning-forming and wild causes, what distinguishes them? In particular, what makes it true that some symbol *F* means [horse] and not [horse or cow on a dark night] or [horse or cow on a dark night or w_2 or w_3 or…] (where each w_i is a property the exemplification of which could cause the tokening of *F*)?

Fodor speculates that a state's being caused by wild conditions depends upon its being caused by meaning-constitutive ones, but not vice versa: ill-lit cows causing horse thoughts depends upon well-lit horses causing them, but well-lit horses causing them does not similarly depend upon the ill-lit cows doing so (getting things wrong depends upon getting things right in a way that getting things right does not depend upon getting things wrong). Fodor draws from this speculation the proposal that mention of ideal conditions is inessential: the structure of this asymmetric causal dependency alone, abstracted from any specific conditions or causal chains, can do all the work.

One problem with this suggestion is that there are many asymmetric dependencies even within meaning-forming cases: for example, small horses causing horse thoughts might asymmetrically depend upon all horses doing so, but not vice versa. If this were so, Fodor would be committed to claiming that 'horse' meant 'non-small horse'.

Another problem is that, in avoiding any mention of mentality, the account risks gratuitous meanings that are brought about by the physics of the world but have no cognitive significance for the agent: electrical stimulation by a poking neurosurgeon, or by cosmic rays, causing tokenings of a mental state in one set of circumstances (for example, when potassium levels were high), might depend upon their lawfully causing those tokenings under other circumstances (when sodium levels were low, for example), but not vice versa. Fodor's view would seem to be committed to treating these further phenomena as meanings, albeit of no cognitive relevance to the agent.

Fodor (1987, 1990a) has replied to these objections with considerable ingenuity. However, what is needed is not only to rule out cases one-by-one, but some general reason to believe that the relevant notion of asymmetric dependency can be explicated without appeal to any mentalistic notions.

Indicator functions. Dretske's most developed ITS proposal is that a mental state has the content 'that *p*' if and only if it is has the *function* of indicating *p*. This distinguishes a proper subset of natural indications; moreover, errors arise when the state fails to indicate what it has the function of indicating. The problem of consequences is handled by observing that from the fact that a state has the function of indicating 'that *p*', and *p* implies *q*, it does not follow that the state has the function of indicating that *q* (see Semantics, teleological; Functional explanation).

The main burden of this approach is to say what makes it the case that a state has a certain indicator function. Dretske offers an account of how a state can acquire such a function, and uses it to offer an account of what he calls 'proto-beliefs' and 'proto-desires'. Suppose that a state *B* of an organism indicates that some property or kind *F* (for example, water) is present in the organism's vicinity, and that a state *D* of the organism renders it receptive to a reward *R*. Then, when in *D*, the organism's behaviour can be reinforced by *R*. Suppose further that when *B* and

D jointly contribute to producing a movement *M* in circumstances in which *F* is present, this results in the organism's receiving *R*. Then, through a process of operant conditioning, *B* and *D* can come to have the 'control' duty of producing *M*. State *B* can thus acquire the control duty because it indicates *F*, and *D* can thus acquire this duty because it is a state of receptivity for *R*. According to Dretske, if *B* and *D* were recruited in this way for the control duty of producing some movement, *B* has the natural function of indicating that *F* is present, and *D* counts as a proto-desire for *R*.

A problem for this account is whether it can be naturalistically supplemented to yield an account of beliefs and desires themselves, since pairs of beliefs and desires do not generally have a specific control duty *vis-à-vis* movements: desiring water and believing there is some in the brook, does not necessarily give rise to any specific behaviour. And, of course, this account is inapplicable for beliefs and desires not acquired by operant conditioning (see Behaviourism, analytic; Behaviourism, methodological and scientific).

Moreover, appeals to function arguably fail to solve the problem of fine-grained individuation. If *p* and *q* are logically equivalent or even nomologically equivalent, then the mechanisms of operant conditioning will be insensitive to their difference (see Intentionality; Propositional attitude statements). Consequently, a state will have the natural function of indicating 'that *p*' if and only if it has the natural function of indicating 'that *q*'. Indeed, whenever *p* and *q* are such that the mechanisms of operant conditioning are insensitive to their difference, whatever has the natural function of indicating *p* has the natural function of indicating *q*. Thus, if, for instance, such mechanisms are insensitive to the difference between the presence of a 'rabbit' and the presence of 'undetached rabbit parts' (to borrow a famous example from Quine 1960), it seems that a state will have the function of indicating that a rabbit is present if and only if it has the function of indicating that undetached rabbit parts are present.

4 Two-factor theories

Since Brentano, it has seemed to many philosophers undeniable that the mind makes more distinctions than even all possible worlds provide: the mind can conceive the impossible, and distinguish among even necessarily co-instantiated properties (like 'being an equiangular triangle' and 'being an equilateral triangle'). 'Two-factor' theories claim that there is something about a mental state in addition to what is captured by ITS that supplies the additional distinctions, notably, its structure or its conceptual role. Fodor (1990a), doubting that conceptual roles can be specified in a way immune to Quine's criticisms of the analytic/synthetic distinction, holds that what distinctions are needed can be captured by structures specified in a language of thought. Others (Block 1986, for example) think that once conceptual roles are specified purely internally, independently of issues of truth, Quine's criticisms are no longer a worry. Most defenders of ITS agree that capturing the contents of logically complex, non-perceptual beliefs (for example, that 'not every particle retains charge if divided') requires reference to some internal factor (see Semantics, conceptual role §3; Concepts §10).

See also: Content, non-conceptual; Information theory; Information theory and epistemology; Semantics; Semantics, teleological

BRIAN P. McLAUGHLIN
GEORGES REY

References and further reading

Block, N. (1986) 'Advertisement for a Semantics for Psychology', in P. French, T. Uehling, Jr., and H. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 10, *Studies in the Philosophy of Mind*, Minneapolis, MN: University of Minnesota Press.(Defence of a 'two-factor' theory of content, allowing for information-theoretic semantics and conceptual role semantics.)

Boghossian, P. (1989) 'The Rule-Following Considerations', *Mind* 98: 507-50. (Raises the holism objection to information-theoretic semantics accounts that appeal to optimal conditions, and argues that ITS accounts cannot capture the normativity of mental content.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press/Bradford Books. (Dretske's original proposal of information-theoretic semantics.)

Dretske, F. (1988) *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press/Bradford Books.(Dretske's later development of information-theoretic semantics.)

Fodor, J. (1984) 'Semantics, Wisconsin Style', *Synthese* 59: 231-50.(Criticism of Stampe- and Dretske-style

theories (see Stampe 1977; Dretske 1981) from a fellow endorser of the information-theoretic semantics strategy.)

Fodor, J. (1987) *Psychosemantics*, Cambridge, MA: MIT Press/Bradford Books. (Source of the 'asymmetric dependency' proposal referred to in §3.)

Fodor, J. (1990a) *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press. (Criticisms of conceptual role theories of content and a defence of his asymmetric dependency proposal.)

Fodor, J. (1990b) 'Psychosemantics, or Where Do Truth Conditions Come From?', in W. Lycan (ed.) *Mind and Cognition: A Reader*, Oxford: Blackwell, 312-38.(A vigorous defence of an 'optimal conditions' version of information-theoretic semantics; widely circulated as a manuscript in 1980 and published in 1990 despite Fodor's repudiation of the view (see Fodor 1987).)

Gates, G. (1996) 'The Price of Information', *Synthese* 107: 325-47.(Forceful criticism of information-theoretic semantics on the basis of the argument of Quine 1960.)

Grice, H. (1957) 'Meaning', *Philosophical Review* 80: 377-88.(Famous effort to distinguish 'natural' from 'non-natural' meaning, aimed against ITS-style theories.)

Loewer, B. (1986) 'From Information to Intentionality', *Synthese* 2: 287-317.(Makes the point about objective probability and natural meaning made in §1; raises the holism objection to accounts using information-theoretic semantics that appeal to optimal conditions.)

Loewer, B. and Rey, G. (eds) (1991) *Meaning in Mind: Fodor and His Critics*, Oxford: Blackwell.(Criticisms by a number of philosophers of Fodor's theory of mind generally, but especially of his 'asymmetric dependency' theory of content; with replies by Fodor and a useful introduction to his theory by the editors.)

McLaughlin, B. (1986) 'What is Wrong With Correlational Psychosemantics', *Synthese* 2: 271-86.(Raises the holism objection to accounts using information-theoretic semantics that appeal to optimal conditions; points out that such accounts cannot cover mental contents that are satisfied by conditions that cannot obtain in optimal conditions, and thus cannot account for beliefs such as that one is not in optimal conditions.)

McLaughlin, B. (ed.) (1991) *Dretske and His Critics*, Oxford: Blackwell.(Contains several essays on Dretske's information semantics and his replies to these essays.)

Peirce, C.S. (1931) *Collected Papers*, ed. C. Hartshorne and P. Weiss, vol. 2, Cambridge, MA: Harvard University Press.(Perhaps the earliest proposal of an information-theoretic semantics.)

Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.(A highly influential discussion of meaning that contains a limited, behaviouristic version of information-theoretic semantics, raising (in chapter 2) the famous 'gavagai' problem for any such theory: how could, for example, 'rabbit' be distinguished from 'undetached rabbit parts'.)

Shannon, C.E. (1948) 'The Mathematical Theory of Communication', repr. in C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Chicago, IL: University of Illinois Press, 1949.(Presents a mathematical theory of *quantity* of information.)

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press/Bradford Books.(Defends a version of information-theoretic semantics that appeals to optimal conditions.)

Stampe, D. (1977) 'Toward a Causal Theory of Linguistic Representation', in P. French, T. Uehling, Jr, and H. Wettstein (eds) *Contemporary Perspectives in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, 81-102. (One of the original proposals of information-theoretic semantics.)

Semantics, possible worlds

Possible worlds semantics (PWS) is a family of ideas and methods that have been used to analyse concepts of philosophical interest. PWS was originally focused on the important concepts of necessity and possibility. Consider:

(a) Necessarily, $2 + 2 = 4$.
(b) Necessarily, Socrates had a snub nose.

Intuitively, (a) is true but (b) is false. There is simply no way that 2 and 2 can add up to anything but 4, so (a) is true. But although Socrates did in fact have a snub nose, it was not necessary that he did; he might have had a nose of some other shape. So (b) is false.

Sentences (a) and (b) exhibit a characteristic known as intensionality*: sentences with the same truth-value are constituent parts of otherwise similar sentences, which nevertheless have different truth-values.* Extensional semantics *assumed that sentences stand for their truth-values, and that what a sentence stands for is a function of what its constituent parts stand for and how they are arranged. Given these assumptions, it is not easy to explain the difference in truth-value between (a) and (b), and hence not easy to give an account of necessity.*

PWS takes a sentence to stand for a function from worlds to truth-values. For each world, the function yields the truth-value the sentence would have if that world were actual. '$2 + 2 = 4$' stands for a function that yields the truth-value 'true' for every world, while 'Socrates had a snub nose' stands for a different function that yields 'true' for some worlds and 'false' for others, depending on what Socrates' nose is like in the world. Since these two sentences stand for different things, sentences that have them as constituents, such as (a) and (b), can also stand for different things.

This basic idea, borrowed from Leibniz and brought into modern logic by Carnap, Kripke and others, has proven extremely fertile. It has been applied to a number of intensional phenomena in addition to necessity and possibility, including conditionals, tense and temporal adverbs, obligation and reports of informational and cognitive content. PWS spurred the development of philosophical logic and led to new applications of logic in computer science and artificial intelligence. It revolutionized the study of the semantics of natural languages. PWS has inspired analyses of many concepts of philosophical importance, and the concept of a possible world has been at the heart of important philosophical systems.

1 Intensions demeaned

Traditionally, the 'intension' of a predicate was distinguished from its 'extension'; the former is a property, the latter is a set. The predicates 'is a featherless biped that is not a plucked chicken' and 'is human' have (one can imagine) the same extensions but different intensions. (The example is from Bertrand Russell.) Gottlob Frege's concepts of *Sinn* and *Bedeutung* extend this idea: the *Sinn* of a singular term is an identifying condition (or 'individual concept'), the *Bedeutung* the individual designated. The *Sinn* of a sentence is a proposition, the *Bedeutung* a truth-value. Frege defended his choice of truth-values as the *Bedeutung* of sentences on systematic grounds (see Frege, G.; Sense and reference).

As model theory was developed by Tarski and others, a version of Frege's choices for *Bedeutung* became the standard values in 'extensional semantics'. The extension of an *n*-place predicate is the set of *n*-tuples of objects of which the predicate is true (thus, the extension of 'gives' might be the set of those four-tuples containing two persons, an object and a time, such that the first person gives the object to the second person at that time). The extension of a singular (object-denoting) term is the object it designates. The extension of a sentence is a truth-value (see Model theory). The packaging together of the predicate calculus with an extensional semantics proved adequate for important work in mathematical logic and overshadowed older approaches to logic. In contrast, no understanding of intensions emerged that is generally agreed on. In the middle part of the century interest in intensional phenomena waned. In fact, the success of extensional logic led to somewhat uncharitable attitudes towards any non-extensional phenomena. In Quine's influential view non-extensional constructions are not suited for scientific work; they are more in need of regimentation than straightforward analysis (see Quine, W.V.). (Non-extensional constructions are those that apparently distinguish between different phrases or sentences

used to complete them, even though the phrases or sentences have the same extensions. One example is 'Elwood believes that…'. 'Elwood believes that Stanford is east of Hawaii' might be true, while 'Elwood believes that Stanford is east of Berkeley' might be false, even though 'Stanford is east of Hawaii' and 'Stanford is east of Berkeley' both have the same extension (the truth-value 'true'). 'Intensional' is sometimes used simply to mean 'non-extensional', and is sometimes given a narrower meaning.)

2 Modal logic

A number of philosophers and logicians continued to attempt to provide straightforward analyses of intensional phenomena, however. Until the 1950s, the emphasis was on syntactic approaches. A key figure was C.I. Lewis, whose dissatisfaction with the extensional treatment of 'if…, then…' as the material conditional led him first to the logic of 'strict implication', then to 'modal logic', the logic of necessity and possibility (see Lewis and Langford 1932). The modal operators (typically translated 'necessarily' and 'possibly', and usually symbolized as $\Box$ and $\Diamond$) are not truth-functional, and so require intensional analysis.

The language of propositional modal logic (ML) consists of the language of propositional logic, plus the rule that if ϕ is a well-formed formula (wff), then so are $\Box\phi$ and $\Diamond\phi$. Lewis and others worked out a number of axiom systems for ML and studied and compared them proof-theoretically.

More semantically oriented approaches to intensionality emerged later in the century, beginning with Carnap (1946, 1947). One of the most important of Carnap's many contributions to the study of intensionality was to recruit Leibniz's idea that necessary truth was truth 'in all possible worlds' to the task of building an intensional semantics. This is the guiding idea of possible worlds semantics (PWS). Carnap's version of this idea, less straightforward than those that were to follow, relies on linguistic representations of possible worlds which he called 'state-descriptions'.

The basics of the now-standard treatment came in the late 1950s and early 1960s with results obtained by Stig Kanger (1957; also Føllesdal 1994) and Saul Kripke (1959, 1963a, 1963b; also Hintikka 1957, Montague 1974a). We shall look briefly at K, S4 and S5, three among the plethora of axiom systems for modal logic that have been studied.

S5 includes:

- all propositional tautologies and *modus ponens*,
- the definition $\Diamond\phi =_{\mathrm{df}} \neg\Box\neg\phi$,
- the rule of necessitation ($\vdash \phi \ / \vdash \Box\phi$, that is, if ϕ is deducible from the null set of premises, then so is $\Box\phi$),
- the axioms

[*K*]: $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$
[*T*]: $\Box\phi \to \phi$
[*4*]: $\Box\phi \to \Box\Box\phi$
[*B*]: $\Diamond\Box\phi \to \phi$

If we drop B we have S4; if in addition we drop *4* and *T* we have **K**, Kripke's minimal system.

A modal model structure is a pair $\langle K, R \rangle$. *K* is the set of worlds; *R* we will consider later. A modal model will tell us which atomic sentences of the base language *L* are true at which worlds of *K*. For the connectives of propositional logic the rules remain unchanged. To extend the system to include $\Box$ it is natural, on the Leibnizian conception, to use the rule (we symbolize ' ϕ is true in *w*' as $\phi[w]$):

$\Box_A$: $\Box\phi[w]$ iff $\forall w' \phi[w']$.

Given the definition of $\Diamond$ we have:

$\Diamond\phi[w]$ iff $\exists w' \phi[w']$.

The reader can check that on this conception, all of the axioms for S5 are valid, that is, true in every world of every model. Consider *B*, for example. Suppose the antecedent is true at *w*. Then $\exists w' \forall w'' \phi[w'']$. Since the existential

quantification is vacuous, this reduces to $\forall w' \phi[w']$. Then, by Universal Instantiation (U.I.), we have $\phi[w]$.

Note that the $\diamondsuit$ was vacuous; $\diamondsuit\Box\phi$ collapsed into $\Box\phi$. This is characteristic of S5: iterated modalities collapse to the right.

S5 is a natural logic for metaphysical necessity, which was doubtless the conception Leibniz had in mind. But there are other coherent concepts of necessity, for which some of the axioms of S5 do not seem correct, and for which distinctions among iterated modalities are significant (see Modal logic).

Consider physical necessity. We have $\Box\phi[w]$ if ϕ is true in every world that obeys the laws of physics of *w*. Suppose w' is a world which has all of our laws and more. Certain events may be ruled out by the physics of w' that are not ruled out by our physics, so our world is not physically possible relative to w'. Suppose, for example, that it is a law of physics in w' that no golf ball travels more than 200 yards. Then even though w' is physically possible (it obeys our laws) and 'No golf ball travels over 200 yards' is necessary in w', it is not true that no golf ball travels over 200 yards. So Axiom *B* is incorrect for physical necessity.

T is intuitive in the case of metaphysical and physical necessity, but not for 'deontic logic', in which ' $\Box\phi$' is interpreted as 'It ought to be the case that ϕ' (see Deontic logic).

In discussing these alternative conceptions of necessity and possibility, we move from an absolute to a relative conception of possibility, the idea that the set of worlds relevant to issues of necessity varies from world to world. This is the information given by the second member of the model structure above. *R* is a relation on *K*, the 'accessibility relation'. Different accessibility relations correspond to different conceptions of necessity. We replace our absolute rule $\Box_A$ with a relative rule:

$\Box_R$: $\Box\phi[w]$ iff $\forall w'$, if w' is accessible from w, then $\phi[w']$.

The axioms that characterize the various systems of modal logic correspond to the logical properties of the relation *R*. The axiom *K* places no restrictions on it; *T* requires reflexivity; *4* requires transitivity; and *B* symmetry. Thus absolute necessity, captured by S5, is the case where the accessibility relation is an equivalence relation.

3 Other applications

The semantical apparatus developed for modal logic has been used to investigate a number of other logical systems.

In 'epistemic logic', for example, a knowledge operator, indexed by knowers, is patterned after $\Box$; $K_\alpha \phi$ means ' ϕ holds in all of α's epistemic alternatives' (Hintikka 1962; see Epistemic logic).

It is important for the philosophically oriented reader to keep in mind that for the purposes of developing and applying semantical treatments of intensional languages, for example, in completeness proofs, the possible worlds of PWS need not be invested with any important metaphysical properties; they are just indices for models. The basic apparatus has been used to study a number of areas in which the metaphor of a possible world is inapplicable. In dynamic logic, for example, the apparatus of modal logic is applied to programs. The 'worlds' are states of a machine. Accessibility relations are indexed by programs. Where α is a program, $\Box[[\alpha]]\phi$ means ' ϕ holds after every terminating execution of α' (see Pratt 1976; Dynamic logics).

The interplay between semantic structures and logical systems involved in these investigations constitute a development in logic comparable to the move in geometry away from Euclidean geometry, conceived as the one true system, to geometry as the study of alternative axiom systems for spaces with diverse properties.

4 Temporal logic

The apparatus of modal model structures works nicely to provide a semantics for temporal logic - the logic of operators modelled after the tense and temporal adverb systems of natural languages. Let '*G* mean 'It will always be the case' and '*F* mean 'It will sometimes be the case'; '*F* can be defined as ' $\neg G \neg$'. Thus *G* is a universal operator, analogous to $\Box$, and *F* is an existential operator, analogous to $\diamondsuit$. Similarly, let '*H* mean 'It has always been the case' and define '*P*' as ' $\neg H \neg$'. Then instead of a set of worlds and an accessibility relation, take a model structure to be a set of moments of time and an ordering relation between them. The need for an accessibility

relation is rather more intuitive here than in the case of necessity and possibility since, unlike worlds, we usually think of times as ordered by the relation of 'before'. As with modal logic, different logics correspond to different conceptions of the ordering relation. One minimal tense logic (Benthem 1988) contains the axioms:

- $G(\phi \rightarrow \psi) \rightarrow (G\phi \rightarrow G\psi)$
- $H(\phi \rightarrow \psi) \rightarrow (H\phi \rightarrow H\psi)$
- $\phi \rightarrow GP\phi$
- $\phi \rightarrow HF\phi$

and the rules *modus ponens* and the analogue to necessitation, sometimes called 'eternity':

- $\vdash \phi / \vdash G\phi$
- $\vdash \phi / \vdash H\phi$

As with modal logic, there is a precise correspondence between ordering conditions and additional axioms. For example $PP\phi \leftrightarrow P\phi$, which seems plausible enough, requires that the structure of moments be transitive (if t is before t' and t' is before t'', t is before t'') and dense (if t is before t', there is a t'' between them, that is, after t and before t') (see Prior 1967; Benthem 1988; Tense and temporal logic).

5 Conditional logic

As we noted, dissatisfaction with the material conditional as an explication of the ordinary-language conditional was an early complaint against extensional logic. There is some connection between the antecedent and the consequent that the semantics for the material conditional misses. For one thing, $\phi \rightarrow \psi$ is true whenever ϕ is false, making all counterfactual conditionals trivially true. Let us use '$\Rightarrow$' as a symbol for a better approximation. Another key way in which $\Rightarrow$ should differ from $\rightarrow$ is that it should not always permit *strengthening the antecedent*. We have:

$(\phi \rightarrow \psi)$ only if $(\phi \wedge \chi \rightarrow \psi)$

but not:

$(\phi \Rightarrow \psi)$ only if $(\phi \wedge \chi \Rightarrow \psi)$.

For example:

ϕ: I put water in my canteen when I start my hike.
ψ: I have water when I stop for a rest.
χ: There is a hole in my canteen.

Robert Stalnaker's version of Frank Ramsey's test for evaluating conditionals is 'make the minimal revision of your stock of beliefs required to assume the antecedent. Then, evaluate the acceptability of the consequent on the basis of this revised body of beliefs' (Ramsey 1931). Stalnaker (1968) and David Lewis (1973) have proposed analyses that implement this idea within the possible worlds framework:

Stalnaker: $\phi \Rightarrow \psi[w]$ iff ψ holds in the ϕ-world that is closest to w.
Lewis: $\phi \Rightarrow \psi[w]$ iff ψ holds in *all* ϕ-worlds which are closest to w.

These analyses require a relation of overall similarity or closeness among worlds. (It can be argued that the relation of overall similarity of worlds is vague and context sensitive; it is replied that this captures the vagueness and context sensitivity of the ordinary conditional.)

On either analysis, strengthening the antecedent fails for $\Rightarrow$, because when the antecedent is strengthened, different worlds might be the closest in which the antecedent is true.

The choice between Lewis' definition and Stalnaker's depends on such issues as whether there is always a unique closest world. One important principle that turns on this is 'conditional excluded middle',

$(\phi \Rightarrow \psi) \vee (\phi \Rightarrow \neg\psi)$

which Stalnaker endorses and Lewis rejects (see Indicative conditionals; Counterfactual conditionals).

6 Quantified modal logic

In 1946, Ruth Barcan Marcus and Rudolf Carnap independently published systems of quantified modal logic (QML), in which principles like the following were considered:

$$\forall x \Box \phi(x) \rightarrow \Box \forall x \phi(x)$$

This is the 'Barcan formula', attractive to those who would reduce *de re* to *de dicto* necessity (see De re/de dicto).

Kripke (1963b) has provided a semantics for these systems. A quantificational modal structure $\langle K, R, \Psi \rangle$ adds a function Ψ which assigns a domain of individuals to each possible world in *K*. A model assigns extensions to each predicate at each world. On the natural, 'world-bound' interpretation of universal quantification, $\forall x \phi(x)$ is true in a world *w* if and only if in *w* $\phi(x)$ is true of every member of the domain of *w*.

If we suppose, as is natural, that worlds have varying domains (that objects that might have been in our world, for example, actually exist in other worlds), then it seems the Barcan formula is not valid. Even if every object we find in the actual world is ϕ in all possible worlds, there could be objects in other possible worlds that are not ϕ. But the formula can be validated if we interpret quantifiers as ranging over all possible objects, or if we suppose that the domain is constant across worlds; each of these alternatives have found proponents.

'Quantifying in' (that is, quantification across modal operators as in the antecedent of the Barcan formula) was deemed by Quine to put us in danger of a commitment to 'essentialism'. Suppose Quine is in the extension of 'is a Kantian' in some possible world *w*. This fact about the realm of possibility is not a fact about the necessary or possible truth of some sentence, so there is some commitment to *de re* modality. Note however that accepting such facts as these does not commit one to a view that Quine must have an essence that distinguishes him from all other objects, one natural understanding of 'essentialism' (Føllesdal 1986) (see Modal logic, philosophical issues in §3).

7 Index theory and intensional logic

Starting in the late 1960s, PWS began to find a role beyond providing semantics for particular logical systems. One development was the development of 'intensional logics' that combined modal, temporal and other operators (see Intensional logics).

We can think of time and possible worlds, for example, as two dimensions along which the truth of a sentence can vary; from this point of view, it is natural to provide a semantics for a system in which sentences are true in worlds at times, and that contains both temporal and modal operators. Thus, for example,

$$\Box(H\phi)$$

will be true in world *w* at time *t* if and only if in every world w' accessible to *w*, at every time t' earlier than *t*, ϕ is true at w' and t'.

But there are also other factors relative to which the truth of sentences can vary. In particular, sentences containing indexicals (such as 'I', 'you', 'here' and 'now') will vary in truth depending on who says them, to whom, where and when.

Thus we could think of:

I will walk to the shop

as being true at a world *w*, time *t* and person *a*, if in *w*, *a* walks to the shop at some time subsequent to *t*.

Following the advice of Dana Scott, Montague, Lewis and others developed and explored versions of 'index theory', systems in which sentences were true at an index, where indices were *n*-tuples of worlds, times, speakers and other factors (see Scott 1970; Montague 1974b; Lewis 1970).

Kamp's 'double-indexing' (1971), Segerberg's 'two-dimensional modal logic' (1973) and Kaplan's 'logic of demonstratives' (1989) provide alternatives to index theory. These authors emphasize the difference between *reliable* sentences such as 'I am here now', that, even if they express contingent propositions, cannot be uttered

falsely, and sentences such as 'either there are cats or there are not', that are valid in the standard sense, of being true in every world in every modal model.

8 Montague semantics

Richard Montague was a leader in the development of intensional logic. In his early papers, he developed PWS as a tool for investigating a number of phenomena of philosophical interest, such as sense-data and events. In later work he developed PWS as a powerful tool for model-theoretic treatments of the semantics of English and other natural languages. Montague's work has had a profound influence in linguistics (see Partee 1989).

As the body of Montague's work developed, intensional phenomena were increasingly seen not as exceptional and marginal, but as at the core of the way language works. In modal logic, intensionality derives from special operators added onto a base language that works on extensional principles. In Montague semantics, intensionality is a ubiquitous phenomenon; there are not only intensional operators, but intensional verbs, adjectives, adverbs and so on. In Montague's later work, intensionality is basically the default case, with special postulates to guarantee extensionality (1974c).

By using intensional logic and PWS to give a precise semantics for constructions of natural language, Montague developed an important new subdiscipline of linguistics, often simply called 'Montague Grammar'.

9 Intensions triumphant

PWS provides philosophy with a toolkit of entities for the analysis of intensional phenomena:

- For individual concepts: functions from worlds to individuals.
- For properties: functions from worlds to extensions.
- For propositions: functions from worlds to truth-values (or sets of worlds).

Note that these functions are themselves extensionally understood, and so analysable within the framework of set theory, and (in that sense at least) free of obscurity.

By the 1970s, philosophers were availing themselves of these tools to talk in disciplined ways about many traditional and some new issues, issues often not directly connected with the interpretation of systems of logic. A few of these are:

- Quantified modal logic has been at the heart of a productive rethinking of issues involved in the distinction between *de dicto* and *de re* necessity, beginning with Quine's charge that quantified modal logic commits us to essentialism (see Føllesdal 1986; Essentialism).
- PWS provides two models for the semantics of names. On the possible worlds version of the Frege-Russell-Searle descriptive account of names, the meaning of a name is an individual concept. An alternative is to model names on variables directly assigned to individuals, the same for all worlds, irrespective of their properties in the worlds. Marcus (1961) suggested the latter possibility, and in *Naming and Necessity* (1980), Saul Kripke has mounted a full-scale challenge to the descriptive account, arguing that names are 'rigid designators' (referring to the same object in all worlds), and providing a causal account of the link between name and thing as an alternative to the descriptive account (see Proper names).
- David Kaplan, whose lectures, seminars and unpublished writings stirred much interest in PWS throughout the 1970s, worked out an account of indexicals and demonstratives in the context of index theory. This has led to a clarification of a number of issues involving the semantics and epistemology of indexicals (see Kaplan 1989; Demonstratives and indexicals).
- David Lewis has used the apparatus of PWS to make significant contributions to our understanding of convention, the semantics of natural language, the understanding of counterfactuals (see above) and many other issues in metaphysics, epistemology and the philosophy of science (see Lewis 1970, 1973, 1979).

Lewis' own view of possible worlds maintains that possible worlds are alternative concrete realities; they are actual for their inhabitants, as ours is for us. The inhabitants of other worlds are not identical with the inhabitants of the actual world (that is, our world), but are their *counterparts*; Lewis' account of quantification is based on the counterpart relations rather than identity.

10 All the intensions we need?

Can PWS supply philosophy with all the intensions that are needed to understand intensional phenomena? Can all intensions be understood as functions from worlds to appropriate sets?

One of the most difficult challenges is the problem of propositional attitudes. The basic problem is that PWS supplies only one necessary proposition (the set of all worlds) and only one contradictory proposition (the null set). This seems to pose a severe problem for dealing with mathematical knowledge. Given usual principles of compositionality, we could infer from 'Elwood knows that $7 + 5 = 12$' to 'Elwood knows that S' for any true mathematical sentence S or any other necessary truth for that matter.

Robert Stalnaker (1984) has given an careful and extended defence of the use of PWS in epistemology. He argues that the concept of content needed for propositional attitudes is grounded in pragmatic relations and such informational relations as indication. The problem of mathematical knowledge, he argues, can be resolved by seeing a linguistic element in our knowledge of mathematical truths.

Advocates of situation semantics have argued, however, that the possible worlds analysis of indication is also vitiated by the problem of the single necessary proposition. Where P is a contingent proposition and N is the necessary proposition, $P = P \,\&\, N$. So if the tree rings indicate that the tree is one hundred years old, they also indicate that it is one hundred years old and $7 + 5 = 12$. But indication appears to distribute over conjunction, so we could infer that the tree rings indicate that $7 + 5 = 12$ (see Perry 1993; Semantics, situation).

While these and other problems have stimulated great interest in other approaches to intensionality in recent years, it seems fair to say that PWS has had by far the most impact on the disciplined investigation of intensional phenomena and that no alternative treatment yet devised provides as natural and comfortable a scheme for thinking about intensional matters.

See also: Intensional entities; Possible worlds

JOHN R. PERRY

References and further reading

Benthem, J. van (1983) *The Logic of Time*, Dordrecht: Reidel.(Important contributions to a wide range of issues in tense logic.)

Benthem, J. van (1988) *A Manual of Intensional Logic*, Stanford, CA: Center for the Study of Language and Information.(A survey of modal logic, tense logic and other applications of the possible worlds framework.)

Carnap, R. (1946) 'Modalities and Quantification', *Journal of Symbolic Logic* 11: 33-64.(Introduces quantifiers into modal logic.)

Carnap, R. (1947) *Meaning and Necessity*, Chicago, IL: University of Chicago Press.(Carnap's classic work; explains his system of intension and extensions, and presents a version of possible worlds semantics.)

Føllesdal, D. (1986) 'Essentialism and Reference', in L.E. Hahn and P.A. Schilpp (eds) *The Philosophy of W.V. Quine*, Library of Living Philosophers, La Salle, IL: Open Court, 97-115.(Discusses Quine's charge that quantification into modal contexts involves essentialism; argues that it is valid in a relatively benign sense of 'essentialism', but not valid in other senses.)

Føllesdal, D. (1994) 'Stig Kanger In Memoriam', in D. Prawitz, B. Skyrms and D. Westerståhl (eds) *Logic, Methodology and Philosophy of Science IX*, Uppsala: Elsevier Science, 885-888.(Discusses the contributions by Kanger, Kripke and others to the semantics of modal logic.)

Hintikka, J. (1957) 'Quantifiers in Deontic Logic', *Societas Scientiarum Fennica, Commentationes humanarum litterarum* 23 (4).(An early use of the idea of relative necessity.)

Hintikka, J. (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Ithaca, NY: Cornell University Press.(Explores the logic of knowledge and belief within a version of the possible worlds framework.)

Hughes, G. and Cresswell, M. (1968) *An Introduction to Modal Logic*, London: Methuen.(Excellent introduction to modal logic and possible worlds semantics.)

Kamp, H. (1971) 'Formal Properties of "Now"', *Theoria* 37: 227-74.(Develops a two-dimensional version of index theory to treat temporal indexicals.)

Kanger, S. (1957) *Provability in Logic*, Stockholm Studies in Philosophy 1, Stockholm: Almqvist & Wiksell.(Basic completeness results in modal logic using accessibility relations.)

Kaplan, D. (1989) *Demonstratives*, in J. Almog, J. Perry and H. Wettstein (eds) *Themes from Kaplan*, 481-563, New York: Oxford University Press.(Classic treatment of indexicals within a possible worlds framework; discussions of related philosophical and semantical issues.)

Kripke, S.A. (1959) 'A Completeness Theorem in Modal Logic', *Journal of Symbolic Logic* 24: 1-15.(Uses possible worlds semantics to prove the completeness of S5.)

Kripke, S.A. (1963a) 'Semantic Analysis of Modal Logic I, Normal Propositional Calculi', *Zeitschrift für mathematische Logic und Grundlagen der Mathematik* 9: 67-96.(Uses possible worlds semantics with accessibility relations to obtain a number of completeness results.)

Kripke, S.A. (1963b) 'Semantic Considerations on Modal Logic', *Acta Philosophical Fennica* 16: 83-94. (Provides possible worlds semantics for quantified modal logic.)

Kripke, S.A. (1965) 'Semantic Analysis of Modal Logic II, Non-normal Propositional Calculi', in I.W. Addison, L. Kenkin and A. Tarski (eds) *The Theory of Models*, Amsterdam: North Holland, 206-20.(Extends semantical treatment to logics without necessitation and other special cases.)

Kripke, S.A. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press. (Groundbreaking work on the semantics of proper names, arguing that they are 'rigid designators'.)

Lewis, C.I. and Langford, C.H. (1932) *Symbolic Logic*, New York: The Century Company.(Includes a survey of modal logics.)

Lewis, D.K. (1969) *Convention*, Cambridge, MA: Harvard University Press.(Develops a conception of convention and argues that language is conventional.)

Lewis, D.K. (1970) 'General Semantics', *Synthèse* 22: 18-67.(Possible worlds semantics in the index style applied to problems in natural language.)

Lewis, D.K. (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press.(Proposes an analysis of conditionals within the possible worlds framework.)

Lewis, D.K. (1979) 'Attitudes *De Dicto* and *De Se*', *Philosophical Review* 88: 513-43. (Treats problems of indexicality and self-knowledge within a possible worlds framework.)

Lewis, D.K. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.(Defends a conception of possible worlds as alternative concrete realities.)

Marcus, R.B. (1946) 'A Functional Calculus of the First Order Based on Strict Implication', *Journal of Symbolic Logic* 11: 1-16.(Introduces quantifiers into modal logic; identifies crucial principles.)

Marcus, R.B. (1961) 'Modalities and Intensional Languages', *Synthèse* 13: 303-32. (Discussion of philosophical issues in the interpretation of quantified modal logic; argues that names are like tags rather than descriptions.)

Montague, R.M. (1974a) 'Logical Necessity, Physical Necessity, Ethics and Quantifiers', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 71-83.(Notes quantificational structure of various operators; uses a relative notion of necessity.)

Montague, R.M. (1974b) 'Pragmatics and Intensional Logic', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 119-147.(Develops an index-style treatment of indexicality within an intensional language.)

Montague, R.M. (1974c) 'English as a Formal Language' and 'The Proper Treatment of Quantification in Ordinary English', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 188-221, 247-70.(In these two classic papers Montague develops a formal semantics for significant fragments of a natural language.)

Montague, R.M. (1974c) 'On the Nature of Certain Philosophical Entities', in R. Thomason (ed.) *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, CT, and London: Yale University Press, 148-87.(Applies possible worlds semantics and other concepts of formal philosophy to sense-data and other classic philosophical problems.)

Partee, B. (1989) 'Possible Worlds in Model-Theoretic Semantics: A Linguistic Perspective', in S. Allen (ed.) *Possible Worlds in Humanities, Arts and Sciences. Proceedings of Nobel Symposium 65*, Berlin and New York: de Gruyter, 93-123.(Survey of the contribution that possible worlds semantics has made to the field of linguistics.)

Perry, J. (1993) *The Problem of the Essential Indexical and Other Essays*, New York: Oxford University Press. (This collection includes papers on indexicality and critiques of possible worlds semantics.)

Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Oxford University Press.(Explores a number of metaphysical issues from the perspective of a version of possible worlds semantics.)

Pratt, V.R. (1976) 'Semantic Considerations on Floyd-Hoare Logic', *Proceedings of the Seventeenth Annual IEEE Symposium on Foundations of Computer Science*, 109-21.(A seminal work in the logic of programs.)

Prior, A.N. (1959) *Time and Modality*, Oxford: Oxford University Press. (Explorations in temporal and modal

logics by a pioneer in the field.)

Prior, A.N. (1967) *Past, Present and Future*, Oxford: Oxford University Press.(The classic presentation of temporal logic.)

Prior, A.N. and Fine, K. (1977) *Worlds, Times and Selves*, London: Duckworth. (Examines metaphysical issues relevant to modal logic and its interpretations.)

Ramsey, F.P. (1931) *The Foundations of Mathematics*, London: Routledge & Kegan Paul.(Collection of Ramsey's philosophical writings.)

Scott, D. (1970) 'Advice on Modal Logic', in K. Lambert (ed.) *Philosophical Problems in Logic*, Dordrecht: Reidel.(An important work in the development of index theory.)

Segerberg, K. (1973) 'Two-Dimensional Modal Logic', *Journal of Philosophical Logic* 2: 77-96.(Develops a version of index theory in which the accessibility relation holds between pairs of indices.)

Stalnaker, R. (1968) 'A Theory of Conditionals', in N. Rescher (ed.) *Studies in Logical Theory*, Oxford: Blackwell.(Proposes an analysis of conditionals within the possible worlds framework.)

Stalnaker, R. (1981) 'Indexical Belief', *Synthese* 49: 129-51.(Introduces the concept of 'diagonal' propositions to handle cases involving indexicals.)

Stalnaker, R. (1984) *Inquiry*, Cambridge, MA: MIT Press.(An analysis of basic epistemological concepts within possible worlds semantics.)

Semantics, situation

Situation semantics attempts to provide systematic and philosophically coherent accounts of the meanings of various constructions that philosophers and linguists find important. It is based on the old idea that sentences stand for facts or something like them. As such, it provides an alternative to extensional semantics, which takes sentences to stand for truth-values, and to possible worlds semantics, which takes them to stand for sets of possible worlds.

Situations are limited parts or aspects of reality, while states of affairs (or infons) are complexes of properties and objects of the sort suitable to constitute a fact. Consider the issue of whether Jackie, a dog, broke her leg at a certain time T. There are two states of affairs or possibilities, that she did or she did not. The situation at T, in the place where Jackie was then, determines which of these states of affairs (infons) is factual *(or* is the case *or* is supported). Situation theory, *the formal theory that underlies situation semantics, focuses on the nature of the* supports *relation.*

Situation semantics sees meaning *as a relation among* types of situations. *The meaning of 'I am sitting next to David', for example, is a relation between types of situations in which someone A utters this sentence referring with the name 'David' to a certain person B, and those in which A is sitting next to B. This relational theory of meaning makes situation semantics well-suited to treat indexicality, tense and other similar phenomena. It has also inspired relational accounts of information and action.*

1 History of situation semantics

Situation semantics was originally conceived as an alternative to extensional model theory and possible world semantics especially suited to the analysis of various problematic constructions, including naked-infinitive perception verbs (Barwise 1981) and belief-reports (Barwise and Perry 1981a) (see Model theory; Semantics, possible worlds). In its earliest forms, the central ideas were:

- *Partiality.* Situations are contrasted with worlds; a world determines the answer to every issue - the truth-value of every proposition. A situation corresponds to the limited parts of reality we in fact perceive, reason about, and live in. What goes on in these situations will determine answers to some issues, but not all. In 'Scenes and Other Situations' (1981), reporting his initial work on situation theory, Jon Barwise represents scenes, the situations we perceive, as partial first-order models.
- *Realism.* Basic properties and relations are taken to be real objects - uniformities across situations and objects - not bits of language, ideas, sets of n-tuples, or functions. In *Situations and Attitudes* (Barwise and Perry 1983), 'courses of events' are partial functions from sequences of locations, relations and objects to truth-values. Complex properties and relations, and various types of objects were full-fledged objects, entering into courses of events.
- *The relational theory of meaning.* The meaning of an expression ϕ is conceived as a relation between a discourse situation, a connective situation and a described situation, written:
$_{d,c}[[\phi]]_e$

The meaning of 'I am sitting next to David', for example, would obtain between courses of events d, c and e if there are individuals a and b such that (1) in d, a is the speaker of the sentence; (2) in c, a's use of 'David' is used to refer to b; and (3) in e, a is sitting next to b.

A number of trenchant criticisms were made of *Situations and Attitudes* (see especially Soames 1985). In reaction to them, Barwise and Perry (1985) recognized the need to rethink the foundations of situation semantics. Two main developments bridge the early versions of situation semantics and the later ones which emerged from this rethinking:

(1) In early versions, situation semantics was developed within standard set theory; this led to foundational problems. In the mid-1980s, Barwise and others developed various versions of 'situation theory', in which all of the various entities that had become necessary were treated axiomatically (Barwise 1989; Devlin 1991; Westerståhl 1990) or within Peter Aczel's version of set theory (Barwise and Etchemendy 1987).

(2) The concept of a constraint, developed in *Situations and Attitudes* as an adjunct to the relational theory of meaning, has become central to the development of situation semantics as a general account of informational and intentional content (Barwise 1993; Israel and Perry 1990; Perry 1993).

2 Situations

The basic idea of situation semantics is that in thought and action we use complexes of objects and properties to classify 'directly' and 'indirectly' parts and aspects of reality, or 'situations'. This sort of realistic classification is more basic than linguistic classification, and it underlies linguistic classification. Consider a simple dialogue:

'What happened in the woods this afternoon?'
'Jackie broke her leg.'

The question concerns a certain situation, a bit of reality: the events in the woods this afternoon. The answer directly classifies the situation in terms of an object (the dog Jackie) and a property (acquiring a broken leg). We classify situations by what goes on in them, by which properties objects have, and by the relations they stand in to one another in virtue of the events that comprise the situation.

Consider the issue of whether Jackie broke her leg at a certain time t. There are two dual possibilities or 'states of affairs', corresponding to whether her leg was broken or not, which we can represent as:

σ: $\langle\langle \text{breaks leg}, t, \text{Jackie}; 1 \rangle\rangle$

and

σ': $\langle\langle \text{breaks leg}, t, \text{Jackie}; 0 \rangle\rangle$

Of course, what goes on in the whole world (if we assume there is such a totality) will determine whether or not Jackie broke her leg, but this will also be determined by much smaller situations. Let s be the situation in the woods this afternoon. Then,

$s \models \sigma$

that is, *s supports* σ, or (in more traditional philosophical terms) *s makes it the case* that σ, or *makes σ factual*.

In situation theory, various objects are built from the basic interplay of situations and states of affairs, permitting complex and abstract ways of classifying situations, including complex states of affairs, properties and relations. A key concept is a 'type of situation', such as the type of situation in which a dog breaks its leg (call it S) and the type of situation in which a dog does not run (call it S').

There are states of affairs involving these abstract objects. In particular, one type of situation may involve another: if there is a situation of the first type, there will also be one of the second type. S involves S': dogs with broken legs do not run. These sorts of states of affairs are 'constraints'.

3 Meaning

Constraints give rise to the possibility of 'indirect classification': classifying situations by what they mean. That is, classifying situations not by the states of affairs they support, but by the types of situations they involve, relative to some constraint.

Indirect classification is how situation semantics conceives of informational and intentional content. Classifying situations by their contents is what organisms do under the influence of what Hume calls 'custom': confronted with a situation, they form expectations, or at least contemplate possibilities, on the basis of what the situation involves relative to some constraint, factual or not, to which they have become attuned. Situation semantics interprets informational and intentional content as a system that exploits such indirect classification. Situations are indirectly classified relative not only to laws of nature and other actual constraints (informational content), but also to conventions, rules, customs, plans and other constraints, both factual and fictional, of human contrivance (intentional content).

Consider:

(1) Jackie has a broken leg.
(2) The x-ray shows that Jackie has a broken leg.
(3) The vet said that Jackie had a broken leg.

In (1), we have direct classification; in (2) and (3), indirect classification. In the latter a common pattern is discernible, involving three types of situation and a constraint:

• A local situation.
• Connections between objects in that situation and other objects, the 'subject matter'.
• A remote situation, the content, involving the subject matter.
• A constraint according to which a combination of situations of the first two types involves a situation of the third type.

In (2) the local situation is the x-ray having certain characteristics. The x-ray is connected to Jackie: it was taken of her. The complex type of situation, in which an x-ray taken of a certain dog exhibits those features, involves a situation in which the dog has a broken leg. Given the connections between the x-ray and Jackie, its having those characteristics shows that she has a broken leg. Here the constraint is factual and the content is informational.

In (3) the local situation is the utterance, in which the vet utters the words 'Jackie has a broken leg'. The vet's use of the word 'Jackie' is connected through various mental and conversational links to the dog Jackie. The rules of English provide the constraint: given the characteristics of the utterance and its connections, it is true if and only if Jackie has a broken leg. English speakers are attuned to these constraints, not in the sense that they automatically form expectations when they hear utterances, but in the sense that they grasp the type of situation meant. Here the content is intentional.

Situation semantics then conceives of meaning as a relation between types of situations. A key advantage of this conception is that it allows us to see how different information can be gleaned from the same 'signal' given different starting points.

In the case of (2), we think naturally of a case in which an experienced vet in a well-organized office studies an x-ray known to be of Jackie, and learns that she has a broken leg. In another case, an experienced vet in a poorly organized office might infer from the x-ray and the fact that Jackie is the only dog in the place with a broken leg, that it was of her. And a would-be vet might learn how to read x-rays, knowing that the x-ray is of Jackie and that she has a broken leg.

Similarly, in (3) we think of a person who knows to which dog the vet refers when he says 'Jackie' and knows English, learning that Jackie has a broken leg. But attunement to the same constraint and a different starting point might allow someone to learn which of the dogs in the office was named 'Jackie'.

4 Accomplishments

Situation semantics has been used to analyse a wide variety of linguistic phenomena (see, for example, Gawron and Peters 1990; Cooper 1992), the liar paradox (Barwise and Etchemendy 1987), heterogeneous reasoning and representation (Barwise and Etchemendy 1991), diagrammatic reasoning (Shin 1990), the nature and structure of information and action (Israel and Perry 1990; Devlin 1991; Barwise 1993), and a number of other issues involving language, representation and computation.

It is probably fair to say, however, that up to this point situation semantics has been more successful in terms of adoption of its broad themes than in terms of adoption of its specific formalism and proposals. The main themes of early situation semantics (partiality, realism and the relational nature of meaning; see §1) have been incorporated into the (generally) received wisdom of philosophy and linguistics. But situation semantics remains only one of a number of alternative semantical frameworks that exhibit these virtues in various ways, from which a theorist may choose.

JOHN R. PERRY

References and further reading

Aczel, P., Israel, D., Katagiri, Y. and Peters, S. (eds) (1993) *Situation Theory and Its Applications*, vol. 3,

Stanford, CA: Center for the Study of Language and Information. (This volume contains a number of articles about situation theory and situation semantics; it is the third in a series of volumes devoted to these topics; see Cooper, Mukai and Perry (1990) and Barwise *et al*. (1991) for the other volumes.)

Barwise, J. (1981) 'Scenes and Other Situations', *Journal of Philosophy* 77: 369-97; repr. in *The Situation In Logic*, Stanford, CA: Center for the Study of Language and Information, 1989.(In this paper Barwise describes his original idea of a scene semantics, a precursor of situation semantics in which partial first-order models are used to provide an analysis of naked-infinitive perception reports.)

Barwise, J. (1989) *The Situation In Logic*, Stanford, CA: Center for the Study of Language and Information.(The papers in the collection developed many of the basic ideas of later versions of situation theory.)

Barwise, J. (1993) 'Constraints, Channels, and the Flow of Information', in P. Aczel, D. Israel, Y. Katagiri and S. Peters (eds) *Situation Theory and Its Applications*, Stanford, CA: Center for the Study of Language and Information, vol. 3, 3-27.(In this paper Barwise develops an abstract theory of information flow based on the idea of constraints.)

Barwise, J. and Etchemendy, J. (1987) *The Liar*, New York: Oxford University Press.(Barwise and Etchemendy provide an analysis of the liar paradox in terms of situation semantics.)

Barwise, J. and Etchemendy, J. (1991) 'Visual Information and Valid Reasoning', in W. Zimmerman (ed.) *Visualization in Mathematics*, Washington, DC: Mathematical Association of America, 9-27.(Barwise and Etchemendy use a situation-based concept of information to defend the view that reasoning using diagrams and reasoning that combines diagrams and sentences can provide valid proofs and can be studied model-theoretically.)

Barwise, J., Gawron, J.M., Plotkin, G. and Tutiya, S. (eds) (1991) *Situation Theory and Its Applications*, vol. 2, Stanford, CA: Center for the Study of Language and Information.(This volume is the second in a series devoted to articles about situation theory and situation semantics; see Cooper, Mukai and Perry (1990) and Aczel *et al*. (1993) for the other volumes.)

Barwise, J. and Perry, J. (1981a) 'Semantic Innocence and Uncompromising Situations', in P. French, T. Uehling and H. Wettstein (eds) *Midwest Studies in the Philosophy of Language*, Minneapolis, MN: University of Minnesota Press, vol. 6, 387-403.(In this article, which announced situation semantics, Barwise and Perry criticize a famous argument used by Church, Davidson and others to show that facts cannot be the semantic values of sentences.)

Barwise, J. and Perry, J. (1981b) 'Situations and Attitudes', *Journal of Philosophy* 77: 668-91.(In this early article Barwise and Perry describe the application of situation semantics to belief-reports and explain the foundational issues that plagued early versions of situation semantics.)

Barwise, J. and Perry, J. (1983) *Situations and Attitudes*, Cambridge, MA: MIT Press/Bradford.(This book developed a version of situation theory within set theory, and applied it to linguistic meaning, the propositional attitudes and the semantics of mental states.)

Barwise, J. and Perry, J. (1985) 'Shifting Situations and Shaken Attitudes', *Linguistics and Philosophy* 8: 105-61.(Responding to the criticisms of Soames and others, Barwise and Perry recognize that their theory (1983) has serious problems and explore strategies for repairing it.)

Cooper, R. (1992) 'A Working Person's Guide to Situation Theory', in S. Hansen and F. Sørensen (eds) *Semantic Representation and Interpretation*, Fredericksberg: Samfundslitteratur.(A lucid introduction to situation theory for the working cognitive scientist.)

Cooper, R., Mukai, K. and Perry, J. (eds) (1990) *Situation Theory and Its Applications*, vol. 1, Stanford, CA: Center for the Study of Language and Information.(This volume is the first in a series devoted to articles about situation theory and situation semantics; see Barwise *et al*. (1991) and Aczel *et al*. (1993) for the other volumes.)

Devlin, K. (1991) *Logic and Information*, Cambridge: Cambridge University Press. (Devlin's book explains the basics of the current version of situation theory and applies it to a number of problems in linguistics and psychology.)

Gawron, J.M. and Peters, S. (1990) *Anaphora and Quantification in Situation Semantics*, Stanford, CA: Center for the Study of Language and Information. (Situation semantics is used to explain the certain aspects of anaphoric relations.)

Israel, D. and Perry, J. (1990) 'What is Information?', in P. Hanson (ed.) *Information, Language, and Cognition*, Vancouver, BC: University of British Columbia Press, 1-19.(Israel and Perry develop a distinction between pure and incremental information within situation theory.)

Israel, D. and Perry, J. (1991) 'Information and Architecture', in J. Barwise, J.M. Gawron, G. Plotkin and S. Tutiya (eds) *Situation Theory and Its Applications*, Stanford, CA: Center for the Study of Language and

Information, vol. 2, 147-60.(Israel and Perry use the theory they developed in 'What is Information?' (1990) to distinguish among various types informational structures.)

Perry, J. (1993) *The Problem of the Essential Indexical and Other Essays*, New York: Oxford University Press.(A collection of Perry's essays on situation semantics and related topics.)

Shin, S.J. (1990) *The Logical Status of Diagrams*, Cambridge: Cambridge University Press.(Situation theory is used to provide a precise analysis of reasoning with Venn diagrams.)

Soames, S. (1985) 'Lost Innocence', *Linguistics and Philosophy* 8: 59-71.(A trenchant critique of Barwise and Perry (1983).)

Westerståhl, D. (1990) 'Parametric Types and Propositions in First-Order Situation Theory', in R. Cooper, K. Mukai and J. Perry (eds) *Situation Theory and Its Applications*, Stanford, CA: Center for the Study of Language and Information, vol. 1, 193-230.(A subtle analysis of the role of parameters in situation theory.)

Semantics, teleological

Teleological/biological theories of meaning use a biological concept of function to explain how the internal states of organisms like ourselves can represent conditions in the world. These theories are controversial, as they have the consequence that an organism's history affects the content of its present thoughts. These theories have advantages over other naturalistic theories of meaning in the task of explaining the possibility of error and unreliable representation.

1 Basic ideas

Teleological/biological theories (hereafter, 'teleological theories') of meaning claim that certain internal states of organisms represent particular external conditions because of the biological *functions* of these inner states and mechanisms that interact with them. These functions derive from evolutionary history, or from some other history of selection; something's function is (roughly) the thing it does which explains why it is there (Wright 1973). For example, the function of the heart is pumping blood, as that is the thing hearts do that explains why they are there. Teleological theories of meaning, such as those theories of Millikan (1984) and Papineau (1987), apply this strategy of analysis to inner representational states (see Functional explanation).

These theories claim that inner states that help an organism coordinate its behaviour with its environment can represent environmental conditions that have played a certain causal role in the organism's history. For example, an animal may be put together in such a way that when a dark shadow appears overhead, an inner state occurs which causes the animal to hide. This inner state has the content 'Predator!' (or perhaps 'Danger!') if it is the presence of predators (or danger) on past occasions of this type that explains this pattern of inner wiring being favoured by natural selection. Similarly, though a frog might snap at any small dark object in its visual field, if the pattern of wiring responsible for this behaviour has been selected over alternatives because of past occasions in which this behaviour resulted in frogs obtaining food, then 'Food!' or perhaps 'Edible insect!' is the content of the inner state that is caused by the dark spot and which causes the behaviour. To use the language of Ruth Millikan (1984), who has developed the most sophisticated teleological theory, the inner state represents the environmental condition that must obtain for the state to help the parts of the organism that make use of (or 'consume') it to perform their functions in a historically normal way.

This type of explanation accounts for the content of belief-like states. The content of desire-like states can be explained in terms of satisfaction conditions. A desire's satisfaction condition is the condition in the world that the desire is supposed to bring about (see Belief; Desire).

Teleological theories do not hold that all beliefs or concepts have a specific evolutionary history. First, these theories can explain some novel beliefs as novel combinations of old concepts. Second, there can be teleological explanations of concepts that are not innate. Millikan explains the semantic properties of learned concepts in terms of the biological functions of parts of the cognitive system that are designed to adapt the organism to novel conditions by means of learning. An alternative approach, defended by Papineau (1987, 1993), is to view learning as a selection process itself, one akin to natural selection across generations, and an independent source of biological functions. In principle there are also other mechanisms which could bestow functions of the relevant sort, such as cultural evolution and deliberate design.

Though Millikan and Papineau give general analyses of all mental representation in these terms, it is also possible to use these ideas in less ambitious ways. A teleological theory could be used to explain only the most basic types of internal representation. Other theories of meaning could then make use of the representational abilities explained by the teleological theory (Sterelny 1990). For example, we might explain a basic stock of primitive concepts in teleological terms, and see more complex concepts as logical constructions out of these (an old-fashioned possibility). Or the teleological view might explain how it is possible for organisms to represent basic environmental features, including other organisms, and then social factors could be used to assign semantic properties to more complex thoughts and public representations. Either way, the teleological view would then be part of a package of different theories which work together in explaining meaning.

2 Related theories

One way to motivate the teleological view is by contrasting it to other naturalistic semantic theories. Indicator theories explain representation in terms of the existence of a reliable correlation, in certain conditions, between an inner state and its object (see Dretske 1981; Fodor 1990; Semantics, informational). These theories have difficulty explaining error and unreliable representation. Teleological theories are attractive to some because they allow an organism to represent something that it cannot, in principle, reliably detect. Suppose an organism's means of detecting predators is not reliable, as it is prone to 'false alarms'. No matter how unreliable it is, if this inner wiring was selected because it enabled organisms to escape predators, the inner state in question can be seen, on a teleological view, as about predators. For an indicator theory, the content of the inner state must be something weaker, which is reliably correlated with the inner state, such as dark shape. Although most teleological theories do assign these 'ecologically salient' states as contents of representations - 'Predator!' rather than 'Dark shape!' - this is not strictly necessary. Neander (1995) has defended a theory which is teleological in structure, but which converges with indicator theories in the contents it assigns in these cases.

Indicator theories are 'upstream-looking' theories; they link a representation to its object by looking to the processes involved in the bringing about of the representational state. They are based upon the organism's powers of discrimination by means of perception. Teleological theories look instead to connections 'downstream' of the representational state, connections going via behaviour and its consequences. In this respect there is a kinship between the teleological approach and some older ideas about thought associated with pragmatism (see Pragmatism). A simple way to link behaviour and belief content is to suppose that the truth condition of a belief is the condition such that actions based on the belief will be successful if and only if that condition obtains. The teleological approach is one way of developing this idea. It replaces the problematic notion of 'success' with a precise concept based upon natural selection. In some respects the debate between the indicator view and the teleological view is an expression of a more general opposition concerning the relative importance of perception and behaviour in understanding the mind. However, there have also been attempts to combine the indicator and the teleological approaches, such as the theory defended by Dretske (1988). For Dretske an inner state represents predators if it has the 'function to indicate' predators; both reliable correlation and function are required.

3 Current debates

Teleological theories of representation have been the subject of lively debate. I will discuss three types of objection.

First, these theories claim that organisms can only represent the world in virtue of facts about their history. Some find this unacceptable, as it entails that a molecule-for-molecule replica of you (such as the 'swampman' of Davidson 1987) which arose instantly by sheer chance would not have thoughts with any semantic content. Millikan's response is simply to embrace this result. Note that while an evolutionary history is required for content on Millikan's view, a more moderate line can be taken by theories such as Papineau's, which treat individual learning as an independent source of functions. Then a swampman would have no thoughts with content initially, but could acquire them within its lifetime.

Second, Jerry Fodor (1990) has argued that teleological theories cannot account for the 'opacity' of attributions of content (see Intentionality; Propositional attitude statements). Fodor claims that if all Fs are Gs in some environment, then anything that was selected in that environment for responding to Fs was selected also for responding to Gs. If so, attributions of content based on biological function cannot distinguish in such cases between representing the presence of an F and representing the presence of a G. Millikan replies that 'selection for' is a causal matter, and hence that attributions of function have as much opacity as causal explanations in general. Then teleological theories will distinguish between representing the presence of an F and representing the presence of a G, when all Fs happen to be Gs. The theories will only fail to distinguish between these two if the properties of being an F and being a G are indistinguishable in their causal powers. Then the question becomes whether a teleological theory ought to distinguish, for example, thoughts about equilateral triangles from thoughts about equiangular ones (see Causation; Property theory).

Lastly, suppose an organism avoids predators by avoiding some totally different environmental condition, such as sunrise, which happens to be reliably correlated with danger from predation (Pietroski 1992). Teleological theories say that the condition in the world with explanatory importance is the object of representation. So an animal that instinctively flees the sun and hence avoids predators is representing to itself 'Predator!' rather than 'Sun!'. This

can be so even if the animal has no ability to detect predators when they are right in front of it. This is a problem for most teleological views, including Millikan's and Papineau's, although not for Dretske's 'mixed' view nor Neander's theory. The problem is hard for standard teleological views because on such views there is not supposed to be a sharp distinction between the object of a representation and the object which explains the representational practice. These theories are explicitly aimed at assimilating the former to the latter.

See also: Evolution, theory of; Functional explanation; Semantics

PETER GODFREY-SMITH

References and further reading

Davidson, D. (1987) 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association* 60: 441-58.(Introduces 'swampman'.)

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.(Classic text for the 'indicator' approach.)

Dretske, F. (1988) *Explaining Behavior*, Cambridge, MA: MIT Press.(Defends a view which combines the teleological approach and the indicator approach.)

Fodor, J.A. (1990) *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.(Attacks the teleological approach and defends an indicator theory.)

Godfrey-Smith, P. (1994) 'A Continuum of Semantic Optimism', in S. Stich and T.A. Warfield (eds) *Mental Representation: A Reader*, Oxford: Blackwell.(Surveys teleological theories, and other theories which use a link between meaning and success.)

Millikan, R.G. (1984) *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.(The most sophisticated teleological theory.)

Millikan, R.G. (1993) *White Queen Psychology, and Other Essays for Alice*, Cambridge, MA: MIT Press. (Various papers defending and developing Millikan's theory; includes 'Biosemantics', the best introduction to her view, and her response to Fodor, 'Speaking up for Darwin'.)

Neander, K. (1995) 'Misrepresenting and Malfunctioning', *Philosophical Studies* 79: 109-41.(Presents a teleological theory and defends it against Fodor, Pietroski and others.)

Papineau, D. (1987) *Reality and Representation*, Oxford: Blackwell.(Defends a teleological theory and discusses connections to debates about realism.)

Papineau, D. (1993) *Philosophical Naturalism*, Oxford: Blackwell.(A further development of Papineau's teleological theory.)

Pietroski, P. (1992) 'Intentionality and Teleological Error', *Pacific Philosophical Quarterly* 73: 267-82.(An interesting criticism of the teleological approach.)

Sterelny, K. (1990) *The Representational Theory of Mind: An Introduction*, Oxford: Blackwell.(A more 'modest' teleological approach.)

Wright, L. (1973) 'Functions', *Philosophical Review* 82: 139-68.(Classic article on functions.)

Semiotics

As the study of signification, semiotics takes as its central task that of describing how one thing can mean another. Alternatively, since this philosophical problem is also a psychological one, its job could be said to be that of describing how one thing can bring something else to mind, how on seeing 'x' someone can be induced to think about 'y' even though 'y' is absent.

A person in whose head 'y' has been brought to mind may be responding to an 'x' someone else has transmitted with the intention of its signifying 'y'; or, mistakenly, responding to an 'x' someone has transmitted in the guileless expectation of its signifying some 'z'; or, often, responding to an 'x' that comes to his notice without anybody's apparent intention at all. Words, for example, generally signify because someone intends them to, and ideally (though not always) they signify what is intended; whereas clouds signify - a coming storm, a whale - because we so interpret them, not because they shaped themselves to convey some meaning.

Obviously the study of signification forms an integral part of the study of thinking, since no object can itself enter the brain, barring fatal mischance, and so it must be represented by some mental (that is, neural) 'x' that signifies it.

Signifiers are equally essential for creatures far lower than humans, as when a chemical signal 'x' emitted by some bacterium signifies to one of its colleagues some 'y' such as 'there's a dearth of food hereabouts'.

There are a number of ways in which an 'x' can signify some 'y', but for humans these are chiefly: by physical association; by physical resemblance; and/or by arbitrary convention.

When we take some 'x' as signifying some 'y' we are often guessing; our guess is subject to checking by interpretative (re)appraisal.

1 The signifier, the signified, the appraisal

In what might be called the standard version of semiotic theory, which in most essentials we owe to the American polymath Charles S. Peirce, our knowledge of the external world, in fact all thinking of any kind, is composed of chains and skeins of linked representations or signifiers ('signs' for short). The world enters our consciousness as highly processed sense-reports, which begin as representations (for instance, the electro-chemical signals into which our light-recepters transform the light striking the retina). So the world is understood, to the extent that it is, via signs, many of which stand for still other signs, each bound to what it signifies by interpretive (re)appraisals, which take account of such things as general knowledge and immediate context. An illustrative example of such a skein of signs is provided by the various ways in which an English word can be encoded. In Morse code the word 'card' is ' $-\cdot-\cdot \cdot - \cdot- \ -\cdot\cdot$', where, to take just the first Morse letter, ' $-\cdot-\cdot$' stands for:

(1) the relatively long and short electrical impulses we call 'dash, dot, dash, dot' (thus sequenced);
(2) the relatively long and short audible beeps we call 'dash, dot, dash, dot' (thus sequenced);
(3) the words 'dash, dot, dash, dot' (thus sequenced);
(4) the Roman letter 'C';
(5) the sound that 'C' spells in the context '-ard', namely /k/.

Each of these stands, in turn, for any or all of the others. Even this is but a bare beginning, since the entire word 'card', however expressed, stands ultimately for a concept, which interpreted or appraised in a particular context might be a playing card, a visiting card, an index card or part of a machine for carding wool.

Though he had many eminent predecessors, it was Peirce who first undertook, in plangent insights scattered over his (mostly unpublished) eighty-odd volumes of papers, to build a comprehensive discipline of semiotics within which all varieties and manners of signification could be given a unified explanation. He entered his final decline with this mighty task unachieved, but he had at least clearly delineated the framework within which almost all semiotic issues, even the semiotic investigations of antiquity, are currently best understood. In particular, it was Peirce who first definitively gave equal triadic footing to the signifier, what it signifies and its interpretative (re)appraisal.

The two greatest semiotic theorists of antiquity were Aristotle and Augustine, with other important contributions from Plato and the Greek Stoics, especially Chrysippus (see Language, ancient philosophy of). Taken as a group, these philosophers distinguished with some care, first, between ’x’ the signifier and ’y’ what it signifies and, second, between the signified ’y’ and, where at issue, the object in the putatively real world that ’y’ represents (Plato, primarily in *Cratylus*; Aristotle, most importantly in *De Interpretatione (Perihermenias)*; Chrysippus; and Augustine in *De Doctrina Christiana*). In addition Aristotle established in the *Prior Analytics* that in the process by which sign is tied to signified, the determination of what the sign signifies is often a probabilistic guess. If I interpret the apparent fact that ’That weathervane is pointing north’ as signifying that ’The local wind is a northerly’, I am jumping from premise to conclusion without invoking the universal premise (’A north-pointing weathervane *always* signifies that the local wind is a northerly’) that would clinch this enthymeme as a syllogism. (Note that here, as so often, ’signifies’ functions like ’implies’.) In fact this missing universal premise is quite properly omitted, since it is false. Weathervanes sometimes mislead. They can be rusted in place; be under the control of mischievous children; be indicating, in the absence of any more recent wind, the direction from which it blew a day ago. In short, once a sign like a weathervane’s pointing north has been guessed to be signifying a current local northerly, there is much room left for an interpretative (re)appraisal of such vital facts as the presence or apparent absence of mischievous children.

Chrysippus nicely demystified the tie between the signifier and what it signifies by arguing that any such tie is made in the mind. (He so emphasized this point, in fact, as to hold that the tie itself is the *only* mental entity, since for him both signifier and signified - his *semenon* and *semenomenon* - were real things in the real world. Whereas nowadays we recognize that it is not the weathervane itself that signifies to us, but our mental image of it. If its image is lacking, the weathervane has not even been seen.)

It was well recognized in antiquity that, while words may be the archetypical signifiers, many signifiers are not words. Aristotle, for instance, mentions in the *Prior Analytics* that a woman’s lactating constitutes a sign (it signifies that she is pregnant); and Augustine, to take another instance, writes in *De Doctrina Christiana* of footprints as signs (of the passage of whoever owned the feet). In this fashion the groundwork was laid for a general science of signs, that is, for semiotics.

Interest in signs and their relations continued during the medieval and early Renaissance periods, notably among the many who seem to have devoted the bulk of their waking hours to writing exegeses of Aristotle. These activities centred, for reasons partly accidental, in Iberia: in Portugal (Fonseca, P. da and John of St Thomas, also known as Jean Poinsot) and, in an earlier and more arcane outbreak, in Catalonia (Llull, R.). Although this work retains much historical interest, it added little to semiotic theory, perhaps because so much of it was devoted to weighing in on one or another of the theological or quasi-theological debates that in those days took the place of free inquiry. With the full Renaissance, however, and the rebirth of interest in everything human, interest in semiotic problems occasionally took a different and more organized turn, notably in the work of George Dalgarno of Scotland (*c.*1619-87).

2 Modern developments

The next important advances came in the late nineteenth and twentieth centuries. First, Charles S. Peirce codified Aristotle’s notion of enthymemic signification into the trivium: ’sign’ (signifier) + signified objective (which Peirce unfortunately shortened to ‘object’) + ’interpretant’. It is worth especial notice that Peirce’s ’interpretant’ plays an essential role in his system, since for him (as for Aristotle) many and perhaps most signs are understood to be tied to their object(ive)s by probabilistic guesswork - by the intuitive leap that Peirce called ’abduction’ - requiring the amplification and partial (de)confirmation available, with luck, from an appraising interpretant. More importantly, in Peirce’s codification each such interpretant is often itself a sign and so in turn is tied to some signified object, or perhaps to more than one such, each being in need of a still further interpretant, and so on, in some cases *ad infinitum*. In this fashion Aristotle’s enthymemic sequence of (minor premise) + (consequent) + (probabilistic substitute for the missing major premise) is converted into a unified chain of signs, each engaged in a triadic structure of sign + object + interpretant.

The three most important theorists since Peirce have been Charles Morris (1901-79), Thomas A. Sebeok (b.1920), and Umberto Eco (b.1932).

Morris' chief contribution was to put living flesh on Peirce's 'interpretant', which for Peirce was a discarnate or at least incorporeal activity - 'pure mind' perhaps - but which for Morris was an interpreter with a brain and a body. This is a gain, since who or what produces and interprets signs has a physical shape that can influence how signs are shaped and related to other signs. (Simple example: the deaf 'speak' their signs with their hands, face and torso; and, compared to vocal signs, the signs of the deaf are predictably different in form, in their modifiability, in their mutual resemblances, and in the ways in which they can mimic what they signify.) Similarly, the primary signs that begin the process of someone's assigning to some sign some significance - the process that is nowadays called 'semiosis' - are for Morris real and observable 'sign-vehicles' of interest in their own right. Both points are important, because how signs are emitted and received by physical entities (such as bodies and brains), and how the signs of a given system (such as the English language) are structured and inter-related, are topics to which Peirce himself gave short shrift. The distinction between corporeal and discarnate semiosis can be summed up in two terms: 'pragmatics' versus 'pure rhetoric', respectively Morris' and Peirce's terms for the study of how signs are related to their interpretants.

Thomas Sebeok's effect on semiotics as a field of study has been pervasive, and is universally recognized. He was co-founder (with Margaret Mead) of the modern semiotic discipline (in 1962); he was founding editor of the defining journal *Semiotica* (from 1968); and he has been the guiding mentor to the entire field. His theoretical contributions are likely to prove just as lasting. He has steadily broadened the scope of semiotic inquiry, for instance by delving into zoösemiotics (the study of infrahumans' signifying activities, in both fact and myth, and of their 'prefigurements of art') and into the sub-area of endosemiotics (the study of signifying among microbiota, neurons included). He has sharpened some of Peirce's most basic notions (see §3) and he has opened the eyes of many to the necessity of defining the logical nature of the basic connective 'signifies' (or 'is a sign of'), as by noting that that connective must be at least potentially reflexive and symmetric.

Umberto Eco, besides having written the 'semiotic' novels that have brought him deserved fame, has been deeply involved in the development of semiotic theory since the late 1960s. He is the author of well-known basic texts, has written on the history of semiotic inquiry and has striven to refine some of Peirce's fundamental concepts.

Among the many other twentieth-century scholars who have also made noteworthy contributions to this burgeoning field are Ferdinand de Saussure, Jacob von Uexküll (1864-1944), Karl Bühler (1879-1963), Roman Jakobson (1896-1982) and Louis Hjelmslev (1899-1965) (see Structuralism in linguistics).

3 Signifying by association, by resemblance, by convention

A given sign may do its signifying in different ways: in the widely used terminology laid down by Peirce, it may signify by physical association with what it signifies (such signifiers are called 'indexes'); by mimicry, that is, by physical resemblance to it ('icons'); and/or arbitrarily and merely by convention ('symbols'). For instance the sketchy representation of a man found on certain urgently sought doors in airports indicates by physical association (location) that this and not the next one is the door to the men's toilet; it indicates by recognizable physical resemblance to a man that this door leads to the men's facility (and not the women's); and it indicates the latter fact partly by convention, since generally it is a stick-figure with a pumpkin head and bears little resemblance to an actual person.

This three-way distinction of types of sign was first drawn by Augustine (*De Doctrina Christiana*), though the distinction between imitative icon and arbitrary symbol, or between symbol and associative index, had been hinted at earlier, respectively by Plato (especially *Cratylus*) and Aristotle (*Prior Analytics*). The distinctions are clear enough in theory, but their application is sometimes aporetic, as when deciding the extent to which a word like 'ding-dong' is imitative (hence 'iconic' in Peirce's sense) or arbitrary (hence 'symbolic'). Indeed, as Plato showed, albeit inadvertently, such discussions are often inconclusive or plain silly. (He argued in *Cratylus* that no Greek noun could be wholly arbitrary, unless imported from Phrygian or some other barbarian language, since any native Greek word could be partitioned, in the crackpot philology of the time, into meaningful subsegments.)

The ternary division of sign-types has a broad acceptance, though alternatives are possible. Certainly a binary division is conceivable - for instance, into Saussure's 'arbitrary' and 'non-arbitrary' types, the latter including both the iconic and the indexical - and it is not difficult to add a fourth or even fifth sign-type. One such is the sign that signifies by modelling the physical process of producing its real object, as the characters of the Korean Hangul

writing-system do (they model the mouth in the act of articulating the characters' corresponding sounds).

On the other hand, Peirce's basic tripartite division of sign-types seems quite robust in comparison with some of the other categorizations that have been proposed. For example, in George Dalgarno's tripartite typology, the earliest fully coherent one (in *Didascalocophus* (1680)), three kinds of signification are defined, but by such a hotchpotch of criteria as to leave the categories leaking rather badly. Dalgarno reserves 'conventional' signs (Peirce's 'symbols') for the exclusive use of humans, and 'natural' signs (Peirce's 'indexes') for infrahumans; supernatural signs, such as 'dreams' and 'apparitions' (these correspond to nothing in Peirce's basic typology) he reserves for the use of 'Almighty God'. (This last category is consistent with the pervasive belief in theurgy typical of Dalgarno's day, but it may have sprung from one of the more bizarre passages in *Cratylus*, where Plato avers that the gods have their own language. Proof: they call 'Xanthus' the river that humans call 'Scamander'.) Dalgarno would thus implicitly deny to humans the ability to signify anything by indexes (such as by correctly placing the 'men's toilet' sign) or, presumably, by such humanly transmitted 'apparitions' as plays or films.

4 Systematicity

Signs typically fall into sets, each sign sharing with its fellows distinctive properties of similarity of form and/or function. Thus the words of English are similar in their partaking of the English sound-system (so that a French word, if properly pronounced, cannot be an English word). Again, any arrangement of signs properly called a sentence is only the overt aspect of one or more covert underlying structures. The sentence: 'Politicians who hold floozies to be congenial are likely to be denounced from the pulpit', can be assigned either of two rather different structures and corresponding meanings. In one, certain politicians embrace floozies in order to be congenial; in the other, it is to the floozies that congeniality is attributed, by certain politicians (see Ambiguity). Such structures have been paid far more attention by linguists than by other semioticists, partly because they are presumably more intricate in language than elsewhere; but, clearly, some other systems of signification have intricacies of their own (think of a map of the London Underground), and much remains to be done.

Another way in which signs are parts of systems is this: the connective 'signifies' has logical or functional properties, which when well defined may be uniform. While the colloquial verb 'signifies' does not map neatly into a single logical connective, any more than the colloquial verb 'implies' does, much is gained by pondering this issue. As a defined propositional connective, 'signifies' is functionally equivalent to 'implies' or even 'if and only if'. For instance, when 'signifies' is narrowly defined then '(p signifies q) & (q signifies r)' signifies 'p signifies r'. (Let 'p' = 'that weathervane points southward', 'q' = 'the local wind is southerly', and 'r' = 'it will soon warm up'.) However, 'signifies' as a nominal connective is rather different. From 'this drawing signifies that cloud' & 'that cloud signifies an impending storm' it certainly does not follow that therefore 'this drawing signifies an impending storm'. Two different meanings of 'signifies' are being played on here - 'represents' and 'betokens' - and mixing them must often produce nonsense.

This said, it remains to observe that the construction of a comprehensive semiotic calculus is very much a work-in-progress.

5 The field

As presently constituted, the field of semiotics exists less as a discipline than as a speciality. The process by which someone receives or emits a sign, and interprets or conceives it - the process generally called 'semiosis' - is investigated by both semioticists and psychologists, mainly in a domain such as the processing of language, whether written or spoken. (In earlier semiotic work issues of semiosis were rather slighted, apart from some inconclusive remarks by Peirce on, for example, the 'firstness' (roughly, accessibility without mediation) of sense-impressions. This is nice, but it sharpens only slightly Augustine's pellucid definition of a sign as 'a thing which, over and above the impression it makes on the senses, causes something else to come into the mind as a consequence of itself' (*De Doctrina Christiana* 2.1.1).) The functioning of signs in literature, music, architecture, painting and drama is mainly pursued by semioticists who are also students of those respective arts. Examples of work in these and other areas of semiotics can be found in the references.

See also: Language, philosophy of; Moscow-tartu school

W.C. WATT

References and further reading

Aristotle (*c.* mid-4th century BC) *The Complete Works of Aristotle*, ed. J. Barnes, Princeton, NJ: Princeton University Press, 1984, 2 vols.(Contains *De Interpretatione*, which discusses signification, among other topics. Also contains *Prior Analytics*)

Augustine (396-426) *De Doctrina Christiana (On Christian Instruction)*, New York: Cima, 1947.(Brilliant and clear-headed discussion of signs and their function.)

Bréhier, E. (1951) *Chrysippe et l'ancien stoïcisme*, Paris: Presses Universitaires de France.(The best introduction to Chrysippus.)

Dalgarno, G. (1680) *Didascalocophus*, Menston: The Scolar Press, 1971.(Odd but fascinating early attempt to categorize signs.)

Deely, J. (1982) *Introducing Semiotic*, Bloomington, IN: Indiana University Press.(Especially good on medieval work.)

Eco, U. (1979) *A Theory of Semiotics*, Bloomington, IN: Indiana University Press.(On semiotics as a study of communicative codes.)

Eschbach, A. and Trabant, J. (eds) (1983) *History of Semiotics*, Philadelphia, PA: John Benjamins.(The study of signs from Plato to Wittgenstein.)

Harris, W.V. (1983) 'On Being Sure of Saussure', *Journal of Aesthetics and Art Criticism* 41: 387-97.(A nice account of Saussure's views on the arbitrariness or conventionality of the signs of language.)

Plato (427-347 BC) *Works*, trans. T. Taylor, London, 1804; repr. New York: Garland, 1984.(Contains *Cratylus*, with notes, which discusses signification, among other topics.)

Posner, R. (1982) *Rational Discourse and Poetic Communication*, Berlin: Mouton.(On the semiotic study of language and literature.)

Ross, W.D. (1965) *Aristotle's Prior and Posterior Analytics*, Oxford: Clarendon Press.(Excellent introduction to Aristotle's insights into signification in relation to logic.)

Sebeok, T.A. (1989) *The Sign and Its Masters*, Lanham, MD: University Press of America.(Authoritative account of modern semiotics.)

Tarasti, E. (1994) *Musical Semiotics*, Bloomington, IN: Indiana University Press.(The standard work in the field.)

Watt, W.C. (ed.) (1994) *Writing Systems and Cognition: Perspectives from Psychology, Physiology, Linguistics, and Semiotics*, Dordrecht: Kluwer Academic Press.(Modern work within the intersection of semiotics and psychology.)

Sense and reference

The 'reference' of an expression is the entity the expression designates or applies to. The 'sense' of an expression is the way in which the expression presents that reference. For example, the ancients used 'the morning star' and 'the evening star' to designate what turned out to be the same heavenly body, the planet Venus. These two expressions have the same reference, but they clearly differ in that each presents that reference in a different way. So, although coreferential, each expression is associated with a different 'sense'. The distinction between sense and reference helps explain the cognitive puzzle posed by identity statements. 'The morning star is the evening star' and 'The morning star is the morning star' are both true, yet the sentences differ in cognitive significance, since the former may be informative, whereas the latter definitely is not. That difference in cognitive significance cannot be explained just by appeal to the references of the terms, for those are the same. It can, however, be naturally accounted for by appeal to a difference in sense. The terms 'the morning star' and 'the evening star' used in the first sentence, having different senses, present the referent in different ways, whereas no such difference occurs in the second sentence.

The distinction between sense and reference applies to all well-formed expressions of a language. It is part of a general theory of meaning that postulates an intermediate level of sense between linguistic terms and the entities the terms stand for. Senses give significance to expressions, which in and of themselves are just noises or marks on a surface, and connect them to the world. It is because linguistic terms have a sense that they can be used to express judgments, to transmit information and to talk about reality.

1 Sense, reference and cognitive significance

The distinction between sense and reference was originally drawn by Gottlob Frege in his 1892 article 'Über Sinn und Bedeutung' ('On Sense and Reference'; see Frege, G. §3). According to Frege, associated with each meaningful expression of a language there is a 'sense' (*Sinn*) that makes the expression significant and determines its 'reference' (*Bedeutung*), that is, the entity the expression applies to or designates. ('*Bedeutung*' is also translated as 'designation', 'designatum', 'denotation' and 'nominatum'. It is also sometimes translated as 'meaning'. This can create confusion, since some of Frege's contemporaries, including Russell, used 'meaning' as the translation of '*Sinn*', most likely on the assumption that the notion corresponding to '*Sinn*' is that of cognitive meaning or significance.) Before 1892, Frege had espoused the view that the significance or content of an expression is given exclusively by that which the expression designates or stands for (what he came to regard as the 'reference'). Observe that on such a view two coreferential terms, such as 'Cicero' and 'Tully', are supposed to have the same content or significance. But if that is so, it seems impossible to explain the obvious difference in cognitive significance between, say, 'Cicero is Cicero', a trivially true and uninformative statement, and 'Cicero is Tully', a potentially informative sentence. The problem of how to explain differences in cognitive significance is not restricted exclusively to statements of identity. 'Cicero was an orator' and 'Tully was an orator' also differ in cognitive significance, since it is possible for a competent speaker to accept one of the statements while rejecting the other. Without postulating a difference in the content or significance of the terms 'Cicero' and 'Tully', it would seem impossible to explain the differences between those two pairs of sentences. The distinction between sense and reference provides for a natural explanation: although the two names are coreferential, they are associated with different senses and those senses present the referent in different ways (see Proper names §1).

2 The sense and reference of different types of expressions

The distinction between sense and reference is intended to apply to all meaningful expressions of a language, whether simple or complex. In the case of definite descriptions it is an extremely natural distinction since it is obvious that different descriptions of one and the same object present that object in different ways. In the case of proper names, the reference is the bearer of the name and the sense is traditionally conceived as some sort of condition satisfied uniquely by the referent. There is no unanimity, however, regarding how to understand such a uniquely identifying condition. Frege suggests (1892) that the sense of a proper name is some descriptive information that applies uniquely to the bearer. Thus, some philosophers interpret Frege's suggestion as entailing the claim than each proper name is synonymous with some definite description. Others, including John Searle (1958), view the sense of a given proper name as a cluster of descriptive information that speakers associate with

the name, and the referent as whatever satisfies most of the attributes in the cluster. Other semanticists do not uphold the descriptive nature of the senses of proper names. For example, David Kaplan (1969) has argued that on Frege's view the sense of a name may be seen as a representation of the referent and so the relation between the sense of a name and its reference is much like the relation between picture and pictured object. Finally, Michael Dummett (1981) is the foremost representative of the view that regards the sense of a proper name, as the sense of any expression, as a mechanism or procedure that, under ideal circumstances, would enable a competent speaker to single out the referent.

The reference of a general term is traditionally taken to be the extension of the term in question. Thus, for example, in the case of a predicate such as 'red' the extension is the set of red things, that is, the set of things to which the predicate correctly applies. And the extension of a relational verb such as 'is taller than' is a set of ordered pairs (since 'being taller than' is a relation between two things) such that the first member of each pair is taller than the second member of the pair. The sense of a general term is a mode of presentation of the set and is responsible for the cognitive content or cognitive significance of the term.

Although it is traditional to represent the reference of a general term as a set, such a choice is strictly incorrect, from the standpoint of pure Fregean orthodoxy. Frege distinguished sharply between concepts and objects. Only the latter are, on his view, 'saturated' entities; entities that subsist on their own. People and artefacts, also numbers and sets, are saturated entities. On the other hand, concepts are 'unsaturated'; they do not have an independent existence, but can only subsist as completed by a saturated entity. On Frege's view, the reference of a general term is a concept, and thus an unsaturated entity. Thus, the reference of a general term is not a set, for the latter is a saturated entity. Still, in the posthumously published comments on 'Über Sinn und Bedeutung' (1892-5), Frege himself stresses that any two general terms refer to the same concept just in case their extension is the same. So, it seems that, even for Frege, equality of extension is a criterion of identity for concepts, so taking the reference of a general term to be a set is not a radical departure from the standard Fregean theory.

Finally, according to the theory, whole sentences express complete thoughts. The 'thought' or sense associated with a sentence is the claim or assertion, traditionally called the 'proposition', expressed by the sentence. In 'Über Sinn und Bedeutung' Frege argues that the sentence's reference is its truth-value. The relation between a sentence and its truth-value should therefore be understood in terms of the relation between a name and its bearer or a definite description and the individual it designates. A true sentence is like a name for 'True', so that all the true sentences (like all the false ones) have one and the same reference; however, since different, non-synonymous sentences express different thoughts, they present that reference in different ways. Frege's argument relies on the assumption (see §4 below) that the reference of a complex expression does not change if components with the same reference are intersubstituted. Thus, for example, if we substitute the coreferential 'the orator who denounced Catiline' for 'Cicero' in the complex expression 'the father of Cicero', the reference of the expression (whoever fathered the famous orator) does not change. In this vein Frege asks what it is that remains constant when we intersubstitute coreferential components in a sentence, for that should be the reference, if anything is. He observes that the thought cannot be the reference, for typically the thought expressed by the sentence is altered: 'Cicero was an orator' and 'The orator who denounced Catiline was an orator' clearly express different thoughts. The truth-value of the sentence is what remains constant in spite of the substitution, so that should be the reference.

3 The multiple roles of sense

The sense of an expression, as we have seen, embodies its cognitive value and is solely responsible for its significance. A term that lacks a sense is therefore not a meaningful constituent of a language; however, an expression may have a sense (that is, be meaningful) and still lack reference. For example, 'the least integer' is a meaningful expression even if there is no object to which it refers. The sense of an expression determines the referent (if the expression has a referent) but, as Frege points out, there is no road back from reference to sense. Aristotle, for example, is the referent of many different terms; there is no mechanism, no procedure that would single out a sense only on the basis of what the reference is.

The sense of a sentence - the thought or proposition it expresses - determines a truth-value as reference. But truth-value typically depends on specific features of the world. Thus, given the way the world is, the sense of the sentence 'Aristotle was a philosopher' determines (we suppose) 'True' as reference. Had Aristotle never thought about philosophical questions, the sense of that sentence would determine 'False' as reference. It is the role of the

sense of a sentence to determine truth-value on whatever conditions obtain, thus the sense accounts for the truth-conditions of a sentence. Because the thought expressed by a sentence is a function of the senses of the component expressions, the sense of an expression is that expression's contribution to the truth-conditions of the sentences in which it figures.

Senses are also what speakers grasp when they understand words. Hence, language learning and linguistic competence are a matter of grasping senses. Moreover, speakers communicate linguistically with one another by grasping the same thoughts. So senses, although mentally graspable entities, should not be confused with private images or ideas. Senses are objective, so that different speakers can grasp the same senses and transmit a given thought to other speakers.

4 Compositionality and substitutivity

The Fregean theory of sense and reference espouses two basic and closely connected semantic principles: the principle of compositionality and the law of substitutivity.

The principle of compositionality states that the reference of a complex expression is a function of the references of its components, plus the way in which those components are combined (see Compositionality). Similarly, the sense of a complex expression is a function of the senses of the component expressions. Thus the reference of the sentence 'The Eiffel Tower is taller than the Empire State Building', a truth-value, is determined exclusively by the references of the meaningful components and their combination. The mode of combination is important, since a different reference would be determined if the components were combined differently, as with the sentence 'The Empire State building is taller than the Eiffel Tower'.

Since, by compositionality, the references of the parts are solely responsible for the reference of the whole, it obviously follows that coreferential terms are intersubstitutable without altering the reference of the complex expressions in which they occur. The latter is known as the law of substitutivity, which states that an expression which figures as part of another expression can be replaced with a coreferential one, and the reference of the complex will remain the same. However, the *sense* of a complex is typically altered when coreferential components are intersubstituted, if they have different senses. Thus, for example, 'Aristotle was Greek' and 'The tutor of Alexander the Great was Greek' express different thoughts but, since the latter sentence results from the exchange of coreferential terms, the reference of the two sentences is the same. (The sense of a complex remains the same if components with the same sense are intersubstituted.)

5 Non-customary contexts

The law of substitution seems to have obvious counterexamples. For example, although 'Cicero' and 'Tully' are coreferential names, the reference of sentences such as 'Mary said that Cicero was a historian' or 'John said, "Cicero was a historian"' may be altered if 'Tully' is substituted for 'Cicero'. But, according to the standard theory of sense and reference, some contexts cause shifts in the sense and reference of the expressions that occur in them. Thus the alleged failure of substitutivity is only apparent, for in the contexts in question the terms being substituted do not have their customary reference. The Fregean theory of sense and reference is essentially contextual: what an expression refers to and what its sense is depend on the context in which the expression occurs.

Consider the sentence 'The word "Cicero" contains six characters'. The reference of a word within quotation marks is not the customary referent, but rather the word itself (see Use/mention distinction and quotation). In our example, the occurrence of 'Cicero' does not refer to Cicero but to the word 'Cicero'. Clearly, the substitution of 'Tully' for 'Cicero' in that context would alter the truth-value of the sentence in question. But since 'Tully' and 'Cicero' are different words, the terms of the attempted substitution are not coreferential in that context, and therefore the observed phenomenon does not violate the principle of substitutivity.

Something similar occurs in the case of indirect quotation, such as 'Mary said that Cicero was a historian' (see Indirect discourse), and belief ascriptions, such as 'John believes that Cicero was a historian' (see Propositional attitude statements §§1-2). Here too the intersubstitution of what seem to be coreferential terms may alter the reference of the sentence. And here again, the breakdown of substitutivity is only apparent, for, according to the theory of sense and reference, expressions in contexts such as these, which Frege called 'oblique', have a non-customary referent. The non-customary referent in these cases - the oblique referent - is the customary sense

of the expression. In oblique contexts the reference of the embedded clause is the thought which that clause customarily expresses, and the reference of each of the clause's component expressions is its customary sense. So, the reference of 'Cicero' in the two sentences considered above is not Cicero but rather the sense 'Cicero' has in customary contexts such as 'Cicero was an orator', 'Cicero was a historian' and so on. In oblique contexts expressions have in turn an oblique sense, whose role is precisely to determine the oblique reference. It is an open question whether the semantic treatment of more complex embeddings requires that expressions have a higher-order sense that determines the oblique sense as reference in the context in question.

See also: Intensional entities §1

GENOVEVA MARTÍ

References and further reading

Burge, T. (1990) 'Frege on Sense and Linguistic Meaning', in D. Bell and N. Cooper (eds) *The Analytic Tradition: Meaning, Thought and Knowledge*, Philosophical Quarterly Monographs series, vol. 1, Oxford and Cambridge, MA: Blackwell, 30-60. (Proposes a nontraditional interpretation of the notion of sense.)

Church, A. (1951) 'A Formulation of the Logic of Sense and Denotation', in P. Henle, H. Kallen and S. Langer (eds) *Structure, Method and Meaning*, New York: Liberal Arts, 3-24.(A formal development of Frege's ideas. Church argues for the need for an infinite hierarchy of senses.)

Donnellan, K.S. (1972) 'Proper Names and Identifying Descriptions', in D. Davidson and G. Harman (eds) *Semantics of Natural Language*, Dordrecht: Reidel, 356-79.(A criticism of the Fregean theory of sense and reference as applied to proper names.)

Dummett, M. (1981) *Frege: Philosophy of Language*, Cambridge, MA: Harvard University Press.(A comprehensive discussion of various aspects of the theory of sense and reference.)

Evans, G. (1982) *The Varieties of Reference*, Oxford: Clarendon Press.(A contemporary Fregean approach to sense and reference. Chapter 1 has a thorough exposition and discussion of the doctrine of sense and reference for proper names.)

Evans, G. (1985) 'Understanding Demonstratives', in *Collected Papers*, Oxford: Oxford University Press. (Applies the theory of sense and reference to indexicals.)

Frege, G. (1879) *Begriffsschrift, eine der arithmetischen nachgebildeten Formelsprachen des reinen Denkens*, Halle: Nebert; trans. 'Conceptual Notation', in *Conceptual Notation and Related Articles*, ed. T. Bynum, Oxford: Clarendon Press, 1972, 101-203.(A sample of Frege's thought before the distinction between sense and reference. §8 gives Frege's first attempted solution to the problem posed by true and informative identity sentences.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. and ed. P.T. Geach and M. Black, 'On Sense and Meaning', in *Translations from the Philosophical Writings of Gottlob Frege*, Oxford: Blackwell, 3rd edn, 1980, 56-78; trans. H. Feigl, 'On Sense and Nominatum', in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Ridgeview, 1982, 85-102.(This is the essay in which Frege introduces the distinction between sense and reference.)

Frege, G. (1892-5) 'Ausführungen über Sinn und Bedeutung', in *Nachgelassene Schriften*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Hamburg: Meiner, 1969-76; trans. P. Long and R. White, 'Comments on Sense and Meaning', in *Posthumous Writings*, ed. H. Hermes, F. Kambartel and F. Kaulbach, Chicago, IL: Chicago University Press, 1979, 118-25.(A posthumously published discussion of some of the major themes of 'Über Sinn und Bedeutung'. Frege focuses specially on the application of the notions of sense and reference to general terms.)

Frege, G. (1918) 'Der Gedanke: eine logische Untersuchung', *Beiträge zur Philosophie des deutschen Idealismus* 1: 58-77; trans. A.M. and M. Quinton, 'The Thought: A Logical Inquiry', *Mind* 65: 289-311; trans. P.T. Geach and R. Stoothoff, 'Thoughts', in *Collected Papers*, ed. B. McGuinness, Oxford: Blackwell, 1984.(Frege's defence of the objectivity of the realm of sense. It also includes a few remarks on the sense of indexicals.)

Furth, M. (1968) 'Two Types of Denotation', *Studies in Logical Theory*, American Philosophical Quarterly Monograph series, vol. 2, Oxford: Blackwell, 9-45. (Discusses unsaturated entities and the reference of general terms.)

Kaplan, D. (1969) 'Quantifying In', in D. Davidson and J. Hintikka (eds) *Words and Objections*, Dordrecht: Reidel, 178-214.(A discussion of Frege's treatment of oblique contexts. Argues for the representational conception of the sense of proper names.)

Parsons, T. (1981) 'Frege's Hierarchies of Indirect Senses and the Paradox of Analysis', in P.A. French, T.E. Uehling, Jr and H.K. Wettstein (eds) *Midwest Studies in Philosophy*, vol. 6, 37-57.(Contends, against Church,

that just one level of oblique sense and reference is sufficient to provide an adequate semantic treatment of all contexts.)

Perry, J. (1977) 'Frege on Demonstratives', *Philosophical Review* 86: 474-97. (Criticizes the theory of sense and reference as applied to indexicals.)

Salmon, N. (1986) *Frege's Puzzle*, Cambridge, MA: MIT Press.(A critical discussion of the puzzle of cognitive significance.)

Searle, J.R. (1958) 'Proper Names', *Mind* 67: 166-73.(On the cluster conception of the sense of proper names.)

Sluga, H. (ed.) (1993) *The Philosophy of Frege*, vol. 4, *Sense and Reference in Frege's Philosophy*, New York: Garland.(Large collection of articles about the topic.)

Speech acts

Making a statement may be the paradigmatic use of language, but there are all sorts of other things we can do with words. We can make requests, ask questions, give orders, make promises, give thanks, offer apologies and so on. Moreover, almost any speech act is really the performance of several acts at once, distinguished by different aspects of the speaker's intention; there is the act of saying something, what one does in saying it, such as requesting or promising, and how one is trying to affect one's audience.

The theory of speech acts is partly taxonomic and partly explanatory. It must systematically classify types of speech acts and the ways in which they can succeed or fail. It must reckon with the fact that the relationship between the words being used and the force of their utterance is often oblique. For example, the sentence 'This is a pig sty' might be used nonliterally to state that a certain room is messy, and further to demand indirectly that it be tidied up. Even when this sentence is used literally and directly, say to describe a certain area of a farmyard, the content of its utterance is not fully determined by its linguistic meaning - in particular, the meaning of the word 'this' does not determine which area is being referred to. A major task for the theory of speech acts is to account for how speakers can succeed in what they do despite the various ways in which linguistic meaning underdetermines use.

In general, speech acts are acts of communication. To communicate is to express a certain attitude, and the type of speech act being performed corresponds to the type of attitude being expressed. For example, a statement expresses a belief, a request expresses a desire, and an apology expresses a regret. As an act of communication, a speech act succeeds if the audience identifies, in accordance with the speaker's intention, the attitude being expressed.

Some speech acts, however, are not primarily acts of communication and have the function not of communicating but of affecting institutional states of affairs. They can do so in either of two ways. Some officially judge something to be the case, and others actually make something the case. Those of the first kind include judges' rulings, referees' decisions and assessors' appraisals, and the latter include sentencing, bequeathing and appointing. Acts of both kinds can be performed only in certain ways under certain circumstances by those in certain institutional or social positions.

1 Levels of speech acts

How language represents the world has long been, and still is, a major concern of philosophers of language. Many thinkers, such as Leibniz, Frege, Russell, the early Wittgenstein and Carnap, have thought that understanding the structure of language could illuminate the nature of reality. However noble their concerns, such philosophers have implicitly assumed, as J.L. AUSTIN complains at the beginning of *How to Do Things with Words* (1962), that 'the business of a (sentence) can only be to "describe" some state of affairs, or to "state some fact", which it must do either truly or falsely'. Austin reminds us that we perform all sorts of 'speech acts' besides making statements, and that there are other ways for them to go wrong or be 'infelicitous' besides not being true. The later Wittgenstein also came to think of language not primarily as a system of representation but as a vehicle for all sorts of social activity. 'Don't ask for the meaning', he admonished, 'ask for the use'. But it was Austin who presented the first systematic account of the use of language. And whereas Wittgenstein could be charged with conflating meaning and use, Austin was careful to separate the two. He distinguished the meaning (and reference) of the words used from the speech acts performed by the speaker using them.

Austin's attention was first attracted to what he called 'explicit performative utterances', in which one uses sentences like 'I nominate', 'You're fired', 'The meeting is adjourned' and 'You are hereby sentenced… ' to perform acts of the very sort named by the verb, such as nominating, firing, adjourning, or sentencing (see Performatives). Austin held that performatives are neither true nor false, unlike what he called 'constatives'. However, he came to realize that constatives work just like performatives. Just as a suggestion or an apology can be made by uttering 'I suggest… ' or 'I apologize… ', so an assertion or a prediction can be made by uttering 'I assert… ' or 'I predict… '. Accordingly, the distinction between constative and performative utterances is, in Austin's general theory of speech acts, superseded by that between saying something and what one does in saying it. This broader distinction applies to both statements and other sorts of speech acts, and takes into account the fact

that one does not have to say 'I suggest… ' to make a suggestion, 'I apologize… ' to make an apology, or 'I assert… ' to make an assertion.

The theory of speech acts aims to do justice to the fact that even though words (phrases, sentences) encode information, people do more things with words than convey information, and that when people do convey information, they often convey more than their words encode. Although the focus of speech act theory has been on utterances, especially those made in conversational and other face-to-face situations, the phrase 'speech act' should be taken as a generic term for any sort of language use, oral or otherwise. Speech acts, whatever the medium of their performance, fall under the broad category of intentional action, with which they share certain general features (see Action; INTENTION §5). An especially pertinent feature is that when one acts intentionally, generally one has a set of nested intentions. For instance, having arrived home without one's keys, one might push a button with the intention not just of pushing the button but of ringing a bell, rousing one's spouse and, ultimately, getting into one's house. The single bodily movement involved in pushing the button comprises a multiplicity of actions, each corresponding to a different one of the nested intentions. Similarly, speech acts are not just acts of producing certain sounds.

Austin identifies three distinct levels of action beyond the act of utterance itself. He distinguished the act *of* saying something, what one does *in* saying it, and what one does *by* saying it, and dubs these the 'locutionary', the 'illocutionary' and the 'perlocutionary' act, respectively. Suppose, for example, that a bartender utters the words, 'The bar will be closed in five minutes.' He is thereby performing the locutionary act of saying that the bar (that is, the one he is tending) will be closed in five minutes (from the time of utterance). Notice that what the bartender is saying, the content of his locutionary act, is not fully determined by the words he is using, for they do not specify the bar in question or the time of the utterance. In saying this, the bartender is performing the illocutionary act of informing the patrons of the bar's imminent closing and perhaps also the act of urging them to order a last drink. Whereas the upshot of these illocutionary acts is understanding on the part of the audience, perlocutionary acts are performed with the intention of producing a further effect. The bartender intends to be performing the perlocutionary acts of causing the patrons to believe that the bar is about to close and of getting them to order one last drink. He is performing all these speech acts, at all three levels, just by uttering certain words.

There seems to be a straightforward relationship in this example between the words uttered ('The bar will be closed in five minutes'), what is thereby said, and the act of informing the patrons that the bar will close in five minutes. Less direct is the connection between the utterance and the act of urging the patrons to order one last drink. Clearly there is no linguistic connection here, for the words make no mention of drinks or of ordering. This indirect connection is inferential. The patrons must infer that the bartender intends to be urging them to leave and, indeed, it seems that the reason his utterance counts as an act of that sort is that he is speaking with this intention. There is a similarly indirect connection when an utterance of 'It's getting cold in here' is made not merely as a statement about the temperature but as a request to close the window or as a proposal to go somewhere warmer. Whether it is intended (and is taken) as a request or as a proposal depends on contextual information that the speaker relies on the audience to rely on. This is true even when the connection between word and deed is more direct than in the above example, for the form of the sentence uttered may fail to determine just which sort of illocutionary act is being performed. Consider, by analogy, the fact that in shaking hands we can, depending on the circumstances, do any one of several different things: introduce ourselves, greet each other, seal a deal, or bid farewell. Similarly, a given sentence can be used in a variety of ways, so that, for example, 'I will call a lawyer' could be used as a prediction, a promise, or a warning. How one intends it determines the sort of act it is.

2 Communicative and conventional speech acts

The examples considered thus far suggest that performing a speech act, in particular an illocutionary act, is a matter of having a certain communicative intention in uttering certain words. Such an act succeeds, and the intention with which it is performed is fulfilled, if the audience recognizes that intention. Not by magic, of course - one must choose one's words in such a way that their utterance makes one's intention recognizable under the circumstances. However, as illustrated above, the utterance need not encode one's intention. So, in general, understanding an utterance is not merely a matter of decoding it.

A specifically communicative intention is a reflexive intention, of the sort characterized by H.P. Grice (1989) (see Communication and intention). This is an intention part of whose content is that it be recognized, indeed be

recognized partly on the basis that this is intended. Accordingly, it is an intention whose fulfilment consists in its recognition. This feature distinguishes acts of communication from most sorts of acts, whose success does not depend on anyone's recognizing the intention with which they are performed. One cannot succeed in running a marathon just by virtue of someone's recognizing one's intention to do so, but one can succeed in stating something, requesting something, and so on, by virtue of one's addressee recognizing that one is stating it, requesting it, or whatever. This is success at the illocutionary level. It is a further matter, a condition on the success of the perlocutionary act, whether the addressee believes what one states or does what one requests.

Now Austin did not take into account the central role of speakers' intentions and hearers' inferences. He supposed that the successful performance of an illocutionary act is a matter of convention, not intention. Indeed, he held that the use of a sentence with a certain illocutionary force is conventional in the peculiar sense that this force can be 'made explicit by the performative formula'. P.F. Strawson (1964) argues that in making this claim Austin was overly impressed by the special case of utterances that affect institutional states of affairs, and should have not taken them as a model of illocutionary acts in general. Austin was especially struck by the character of explicit performative utterances, in which one uses a verb that names the very type of act one is performing. For them he developed an account of what it takes for such acts to be performed successfully and felicitously, classifying the various things that can go wrong as 'flaws', 'hitches' and other sorts of 'infelicities'. It is only in certain conventionally designated circumstances and by people in certain positions that certain utterances can have the force they do. For example, only in certain circumstances does a jury foreman's pronouncement of 'Guilty' or 'Not guilty' count as a verdict, a legislator's 'Aye' or 'Nay' as a vote, and a baseball umpire's cry of 'Y'er out' as calling a runner out. In these cases it is only by conforming to a convention that an utterance of a certain form counts as the performance of an act of a certain sort. However, as Strawson argues, most illocutionary acts succeed not by conformity to convention but by recognition of intention. They are not conventional except in the irrelevant sense that the words and sentences being used have their linguistic meanings by virtue of convention (see Language, conventionality of).

Strawson's argument raises a serious problem for theories inspired by Austin's view. Consider, for example, the theory advanced by John Searle (1969), who proposes to explain illocutionary forces by means of 'constitutive rules' (conventions) for using 'force-indicating' devices, such as performative verbs and sentential moods. The problem is that the same sorts of illocutionary acts that can be performed by means of such devices can be performed without them. For example, one does not have to use a performative, as in 'I demand that you be quiet', or the imperative mood, as in 'Be quiet!', to demand that someone be quiet. Clearly a theory that relies on rules for using such devices is not equipped to explain the illocutionary forces of utterances lacking such devices. No such difficulty arises for a theory according to which most illocutionary acts are performed not with an intention to conform to a convention but with a communicative intention.

3 Types of speech acts

Pre-theoretically, we think of an act of communication, linguistic or otherwise, as an act of expressing oneself. This rather vague idea can be made more precise if we get more specific about what is being expressed. Take the case of an apology. If you utter '(I'm) sorry I didn't call back' and intend this as an apology, you are expressing regret for not returning a phone call. An apology just *is* the act of (verbally) expressing regret for, and thereby acknowledging, something one did that might have harmed the hearer. An apology is communicative because it is intended to be taken as expressing a certain attitude, in this case regret. It succeeds as such if it is so taken. In general, an act of communication succeeds if it is taken as intended. That is, it must be understood or, in Austin's words, 'produce uptake'. With an apology, this is a matter of the addressee recognizing the speaker's intention to be expressing regret for some deed or omission. Using a special device such as the performative 'I apologize' may of course facilitate understanding (understanding is correlative with communicating), but in general this is unnecessary. Communicative success is achieved if the speaker chooses their words in such a way that the hearer will, under the circumstances of utterance, recognize their communicative intention. So, for example, if you spill some beer on someone and say 'Oops' in the right way, your utterance will be taken as an apology.

In saying something one generally intends more than just to communicate - getting oneself understood is intended to produce some effect on the listener. However, our speech act vocabulary can obscure this fact. When one apologizes, for example, one may intend not merely to express regret but also to seek forgiveness. Seeking

forgiveness is, strictly speaking, distinct from apologizing, even though one utterance is the performance of an act of both types. As an apology, the utterance succeeds if taken as expressing regret for the deed in question: as an act of seeking forgiveness, it succeeds if forgiveness is thereby obtained. Speech acts, being perlocutionary as well as illocutionary, generally have some ulterior purpose, but they are distinguished primarily by their illocutionary type, such as asserting, requesting, promising and apologizing, which in turn is distinguished by the type of attitude expressed. The perlocutionary act is a matter of trying to get the hearer to form some correlative attitude and in some cases to act in a certain way. For example, a statement expresses a belief and normally has the further purpose of getting the addressee to form the same belief. A request expresses a desire for the addressee to do a certain thing and normally aims for the addressee to intend to and, indeed, actually do that thing. A promise expresses the speaker's firm intention to do something, together with the belief that by their utterance they are obliged to do it, and normally aims further for the addressee to expect, and to feel entitled to expect, them to do it.

Statements, requests, promises and apologies are examples of the four major categories of communicative illocutionary acts: *constatives*, *directives*, *commissives* and *acknowledgements*. This is the nomenclature used by Bach and Harnish (1979) who develop a detailed taxonomy in which each type of illocutionary act is individuated by the type of attitude expressed (in some cases there are constraints on the content as well). There is no generally accepted terminology here, and Bach and Harnish borrow the terms 'constative' and 'commissive' from Austin and 'directive' from Searle. They adopt the term 'acknowledgement', over Austin's 'behabitive' and Searle's 'expressive', for apologies, greetings, congratulations, and so on, which express an attitude regarding the hearer occasioned by some event that is thereby being acknowledged, often in satisfaction of a social expectation. Here are assorted examples of each type:

Constatives. Affirming, alleging, announcing, answering, attributing, claiming, classifying, concurring, confirming, conjecturing, denying, disagreeing, disclosing, disputing, identifying, informing, insisting, predicting, ranking, reporting, stating, stipulating.
Directives. Advising, admonishing, asking, begging, dismissing, excusing, forbidding, instructing, ordering, permitting, requesting, requiring, suggesting, urging, warning.
Commissives. Agreeing, guaranteeing, inviting, offering, promising, swearing, volunteering.
Acknowledgements. Apologizing, condoling, congratulating, greeting, thanking, accepting (acknowledging an acknowledgement).

Bach and Harnish spell out the correlation between type of illocutionary act and type of expressed attitude. In many cases, such as answering, disputing, excusing and agreeing, as well as all types of acknowledgement, the act and the attitude it expresses presuppose a specific conversational or other social circumstance.

For types of acts that are distinguished by the type of attitude expressed, there is no need to invoke the notion of convention to explain how it can succeed. The act can succeed if the hearer recognizes the attitude being expressed, such as a belief in the case of a statement and a desire in the case of a request. Any further effect it has on the hearer, such as being believed or being complied with, or just being taken as sincere, is not essential to its being a statement or a request. Thus an utterance can succeed as an act of communication even if the speaker does not possess the attitude they are expressing: communication is one thing, sincerity another. Communicating is as it were just putting an attitude on the table; sincerity is actually possessing the attitude one is expressing. Correlatively, the hearer can understand the utterance without regarding it as sincere - for example, take it as an apology, as expressing regret for something, without believing that the speaker regrets having done the deed in question. Getting one's audience to believe that one actually possesses the attitude one is expressing is not an illocutionary but a perlocutionary act.

4 Direct, indirect and nonliteral speech acts

As Austin observed, the content of a locutionary act (what is said) is not always determined by what is meant by the sentence being uttered. Ambiguous words or phrases need to be disambiguated (see Ambiguity) and the references of indexical and other context-sensitive expressions need to be fixed in order for what is said to be determined fully (see Demonstratives and Indexicals). Moreover, what is said does not determine the illocutionary act(s) being performed. We can perform a speech act (1) directly or indirectly, by way of performing another

speech act, (2) literally or nonliterally, depending on how we are using our words, and (3) explicitly or inexplicitly, depending on whether we fully spell out what we mean.

These three contrasts are distinct and should not be confused. The first two concern the relation between the utterance and the speech act(s) thereby performed. In indirection a single utterance is the performance of one illocutionary act by way of performing another. For example, we can make a request or give permission by way of making a statement, say by uttering 'I am getting thirsty' or 'It doesn't matter to me', and we can make a statement or give an order by way of asking a question, such as 'Will the sun rise tomorrow?' or 'Can you clean up your room?' When an illocutionary act is performed indirectly, it is performed by way of performing some other one directly. In the case of nonliteral utterances, we do not mean what our words mean but something else instead. With nonliterality the illocutionary act we are performing is not the one that would be predicted just from the meanings of the words being used, as with likely utterances of 'My mind got derailed' or 'You can stick that in your ear'. Occasionally utterances are both nonliteral and indirect. For example, you might utter 'I love the sound of your voice' to tell someone nonliterally (ironically) that you cannot stand the sound of their voice and thereby indirectly to ask them to stop singing.

Nonliterality and indirection are the two main ways in which the semantic content of a sentence can fail to determine the full force and content of the illocutionary act being performed in using the sentence. They rely on the same sorts of processes that Grice discovered in connection with what he called 'conversational implicature' (see Implicature), which, as is clear from Grice's examples, is nothing more than the special case of nonliteral or indirect constatives made with the use of indicative sentences. A few of Grice's examples illustrate nonliterality, such as 'He was a little intoxicated', used to explain why a man smashed some furniture, but most of them are indirect statements, such as 'There is a garage around the corner', used to tell someone where to get petrol, and 'Mr X's command of English is excellent, and his attendance has been regular', giving the high points in a letter of recommendation. These are all examples in which what is meant is not determined by what is said. However, Grice overlooks a different kind of case, marked by contrast (3) listed above.

There are many sentences whose standard uses are not strictly determined by their meanings but are not implicatures or figurative uses either. For example, if one's spouse says 'I will be home later', they are likely to mean that they will be home later that night, not merely some time in the future. In such cases what one means is an *expansion* of what one says, in that adding more words ('tonight', in the example) would have made what was meant fully explicit. In other cases, such as 'Jack is ready' and 'Jill is late', the sentence does not express a complete proposition. There must be something which Jack is being claimed to be ready for and something which Jill is being claimed to be late for. In these cases what one means is a *completion* of what one says. In neither sort of case is a particular word or phrase being used nonliterally and there is no indirection. They both exemplify what may be called 'impliciture', since part of what is meant is communicated not explicitly but implicitly, by way of expansion or completion.

5 Philosophical importance of speech act theory

The theory of speech acts has applications to philosophy in general, but these can only be illustrated here. In ethics, for example, it has been supposed that sentences containing words like 'good' and 'right' are used not to describe but to commend, hence that such sentences are not used to make statements and that questions of value and morals are not matters of fact. This line of argument is fallacious. Sentences used for ethical evaluation, such as 'Loyalty is good' and 'Abortion is wrong', are no different in form from other indicative sentences. Whatever the status of their contents, they are standardly used to make statements. This leaves open the possibility that there is something fundamentally problematic about their contents. Perhaps such statements are factually defective and, despite syntactic appearances, are neither true nor false. However, this is a metaphysical issue about the status of the properties to which ethical predicates purport to refer. It is not the business of the philosophy of language to determine whether or not there are such properties as goodness or rightness and whether or not the goodness of loyalty and the rightness of abortion are matters of fact. The above argument is but one illustration of what Searle calls the 'speech act fallacy'. He also identifies examples of the 'assertion fallacy', whereby conditions of making an assertion are confused with what is asserted. For example, one might fallaciously argue, on the grounds that because one would not assert that one believes something if one was prepared to assert that one knows it, that knowing does not entail believing. Grice identifies the same fallacy in a parallel argument, according to which

seeming to have a certain feature entails not actually having that feature (see Ordinary language philosophy).

For philosophy of language in particular, the theory of speech acts underscores the importance of the distinction between language use and linguistic meaning (see Pragmatics; Semantics). This distinction sharpens the formulation of questions about the nature of linguistic knowledge, by separating questions about capacities exercised in linguistic interaction from those specific to knowledge of language itself. A parallel distinction, between speaker reference and linguistic reference (see Reference), provokes the question of to what extent linguistic expressions refer independently of speakers' use of them to refer. It is common, for example, for philosophers to describe expressions like 'the car', 'Robert Jones' and 'they' as having different references in different contexts, but it is arguable that this is merely a misleading way of saying that speakers use such expressions to refer to different things in different contexts.

See also: Grice, H.P.; Language, philosophy of; Pragmatics; Semantics

KENT BACH

References and further reading

Austin, J.L. (1962) *How to Do Things with Words*, Cambridge, MA: Harvard University Press.(Develops the distinction between performative and constative utterances into the first systematic account of speech acts.)

Bach, K. (1994) 'Conversational Implicature', *Mind and Language* 9: 124-62.(Identifies the middle ground between explicit utterances and Gricean implicatures.)

Bach, K. and Harnish, R.M. (1979) *Linguistic Communication and Speech Acts*, Cambridge, MA: MIT Press. (Combines elements of Austin's taxonomy and Grice's theory of conversation into a systematic account of the roles of the speaker's communicative intention and the hearer's inference in literal, nonliteral and indirect uses of sentences to perform speech acts.)

Grice, H.P. (1989) *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.(The essays on meaning and conversational implicature provide a framework for distinguishing speaker meaning from linguistic meaning and for explaining their relationship.)

Searle, J.R. (1969) *Speech Acts: An Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.(Presents a theory of speech acts relying on the notion of constitutive rules.)

Strawson, P.F. (1964) 'Intention and Convention in Speech Acts', *Philosophical Review* 73: 439-60.(Applies Grice's account of meaning to support the claim that most speech acts are communicative rather than conventional, as Austin had suggested.)

Tsohatzidis, S.L. (ed.) (1994) *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, London: Routledge.(Collection of original essays on outstanding problems in the field, with useful bibliography.)

Structuralism in linguistics

The term structural linguistics can be used to refer to two movements which developed independently of each other. The first is European and can be characterized as post-Saussurean, since Saussure is generally regarded as its inspiration. The central claim of this movement is that terms of a language of all kinds (sounds, words, meanings) present themselves in Saussure's phrase 'as a system', and can only be identified by describing their relations to other terms of the same language; one cannot first identify the terms of a language and then ask which system they belong to. Moreover, because a language is a system of signs, one cannot identify expression-elements (sounds, words) independently of the content-elements (meanings), so that a study of language cannot be divorced from one of meaning. The second movement is an American one, which developed from the work of Leonard Bloomfield and dominated American linguistics in the 1940s and 1950s. It attached great importance to methodological rigour and, influenced by behaviourist psychology, was hostile to mentalism (any theory which posits an independent category of mental events and processes). As a result, unlike the first movement, it excluded the study of meaning from that of grammar, and tried to develop a methodology to describe any corpus in terms of the distribution of its expression-elements relative to each other. Whereas the first movement provided a model for structuralist thought in general, and had a significant impact on such thinkers as Barthes, Lacan and Lévi-Strauss, the second made a major contribution to the development of formal models of language, however inadequate they may seem now in the light of Chomsky's criticisms.

1 The concept of structural linguistics

Though a wide range of linguistic studies raise issues which are in some sense structural, we shall focus on two uses of the term *structural linguistics* which refer to separate movements, each of which advanced strong hypotheses about the nature of language and the methodology of linguistics. However, though each of the movements attaches great importance to structural analysis, their reasons for doing so differ profoundly; hence, the need to distinguish them.

The origin of the first movement is to be found in the work of Ferdinand de Saussure, the fundamental contention of which is that, since linguistic items of all kinds (sounds, words, meanings) present themselves as a system, they can only be identified in terms of the relations in which they stand to other elements within the system (Lyons 1973: 6). An acceptance of this thesis, together with one, sometimes qualified, of the following three Saussurean theses, defines post-Saussurean structuralism: it is necessary to distinguish *la langue* from *la parole*, and a synchronic study from a diachronic one; the fundamental linguistic study is a synchronic one of *la langue*; and, finally, a study of expression-elements (sounds and words) cannot be divorced from one of meanings. Of course, within this framework much is underdetermined, leaving scope for different emphases and developments, two of which are considered in §§2 and 3 below.

The origins of the second movement are to be found in the work of the American linguist Leonard Bloomfield, and, in particular, in his book *Language* (1933). Strongly influenced by behaviourist psychology (see Behaviourism, methodological and scientific), Bloomfield proposed an analysis of a speech act in terms of the notions of a *stimulus* and a *response*. In response to a feeling of hunger (a stimulus) one thing I might do is reach for an apple; but if I have mastered a language, another thing that I can do is to utter 'Pass me the apple', which is itself not only a response to my hunger but also a stimulus designed to get you to respond by passing the apple. So language increases the range of responses enormously by bridging the gap between different nervous systems. A problem with this analysis, as Bloomfield concedes, is that in most cases we seem unable to identify independently either the stimulus or the response. Since he defines the meaning of an utterance as a function of the situation in which it is uttered and the response called forth, it is hardly surprising that he concludes that the study of meaning largely evades our grasp. What can be studied nevertheless are the utterances themselves but, to avoid a lapse into a discredited mentalism, or an appeal to a theory which is not available to us, sound methodology requires that the study of language does not appeal to meaning. The study of language becomes, therefore, a study of a set of utterances, or of a corpus - indeed, Bloomfield once defined a language as the set of utterances uttered by its speakers. Such a study concentrates on the formal relations between constituents of utterances. It argues that for each constituent there is a limited range of environments in which it can occur; for instance, 'cat' can occur in the environment 'The   jumped', but not in 'John   Jill'. So if an item's distribution is defined as the ranges of

environments in which it can occur nondeviantly, there seem to be promising ways of defining its type in terms of its distribution, and of making generalizations about the relations between types. More and more sophisticated approaches to the analysis of a corpus in terms of the distributional relations holding between its constituents were developed in America in the 1940s and 1950s, culminating in the work of Z.S. Harris which will be discussed in §4. So three main features of post-Bloomfieldian structuralism are: it was corpus based; an appeal to meaning was rigorously excluded from the methodology adopted; and the object of analysis was the identification of the constituents of the utterances in a corpus, and the construction of a compact description of the distributional relations holding between them. Paradoxically, whilst it is clear why both movements should be regarded as structuralist, each from the point of view of the other is radically misconceived. Arguably the first movement is the more interesting from a philosophical point of view because of its impact on structuralist thought in general; but the second movement undoubtedly developed much more detailed formal models than the first.

2 The Prague School and functionalism

The central thesis of post-Saussurean structuralism is that the terms of a language present themselves as a system; but precisely what kind of a system do they belong to? A group of linguists whose work was strongly associated with Prague and who became known as the Prague School developed a distinctive answer to this question. For them a language is a dynamic system with many subsystems which interact with each other; one of these concerns the structure of its functional sounds (phonology), another the structure of its words (morphology), another the structure of its sentences (syntax), and so on. Each of these has a stable core, surrounded by a less stable periphery, and is striving for a state of equilibrium which it never quite attains, since a change which stabilizes one part of the system can adversely affect another. The Russian linguist Jakobson, who played a major part in the development of this conception, likened the attempt to restore equilibrium to the moves made in a chess game after the loss of a piece, which may strengthen one part of a defence at the expense of another. The Prague School's dynamic conception of a language as a system of subsystems striving to reach equilibrium led it in turn to adopt a different conception of diachrony from that held by Saussure, for all that it agreed with him that a synchronic perspective is logically prior. Since a language is a dynamic system, the School argued that the force for change in many cases is a state of the system itself, whereas Saussure had argued that change is brought about by factors external to the system which never affect it directly.

An important area in which the School made a major contribution to our understanding of what the system involved was phonology, the study of the functional sound units of a language. It seems clear that all languages utilize a relatively small number of such units (called phonemes), so that 'cat', for instance, contains /c/, /a/ and /t/; but how are they to be characterized, and in what way do they form a system? Saussure was prevented from answering this question in part because he assumed that phonemes were not complex, and that oppositions arise from differences. However, Trubetzkoy and Jakobson questioned both of these assumptions. They argued that phonemes can be analysed in terms of so called distinctive features, which are the ways in which phonemes belonging to the same language are opposed to each other, and that the differences between them arise from these oppositions. For instance, one way in which /m/ differs from /n/ is that the lips are closed when pronouncing it, so that there is a feature (labial) present in one case which is absent in the other. Features of this kind the School called 'distinctive features', claiming that each phoneme can be described in terms of the presence or absence of a small number of such features. Moreover, since what is being described is a set of ways in which the phonemes of a language can be opposed to each other, differences rest on oppositions. This theory was to have a major impact on structuralist thinkers; for example, Lévi-Strauss asked whether there are other kinds of oppositions which give rise to different kinds of structures in domains such as kinship and myth (see Structuralism in social science).

Another novel feature of the Prague School was its attempt to combine a functionalist conception of language with its structuralist conception. The central idea is that a study of a language cannot be divorced from its uses: 'language is not a self-contained whole, hermetically separated from the extra-lingual reality, but in fact its main function is to react to and to refer to this reality' (Vahek 1966: 7). Just as one cannot understand chess fully unless one understands what it is players are trying to achieve, so one cannot understand a language if one does not understand its uses. Moreover, various aspects of use have specific linguistic means of expression; for example, it is natural if one wants someone to do something to use an imperative, if one wants information to use an interrogative, and so forth. Whether functionalism is ultimately consistent with the strong form of the principle of the arbitrariness of the sign embraced by Saussure is questionable. However, by recognizing the diverse uses of

language, the Prague School greatly extended the range of domains open to structural analyses. For instance, if an utterance may have an expressive function as well as a descriptive one, may we not, as a limiting case, have a text which is purely expressive but which can nevertheless be analysed structurally - a poem for instance?

3 The Copenhagen School and glossematics

Arguably, the Prague School's theory of oppositions, which positively characterizes one item in an opposition, is incompatible with the Saussurean theses that in language there are only differences without positive terms, and that language is a form not a substance. The work of the Danish linguist Hjelmslev, which takes these theses as its cornerstone, is in this respect a more faithful development of the Saussurean conception of a system.

The fact that a given domain is structured in different ways by different languages shows, Hjelmslev argues, that a language is a form not a substance, with its own autonomous principles of organization. In the domain of colour, for instance, we find that 'Welsh lacks the English boundary between *blue* and *grey*, and likewise the English boundary between *grey* and *brown*' (Hjelmslev 1943: 53); other examples of this phenomenon are systems of tense and pronouns. But how can an autonomous system be described without characterizing some elements positively, and so appealing to something external to it? This can be done, Hjelmslev argues, by characterizing the elements *only* in terms of the relations they can stand in to other items, so that it matters not at all what the elements are, provided that *they are different from each other*.

The resulting description is a linguistic algebra which Hjelmslev called a *linguistic schema*; whilst the result of interpreting a schema, that is of applying it to a domain, he called a *linguistic usage*. The difference between schema and usage corresponds to a more and a less abstract conception of *la langue*; and for Hjelmslev the primary object of linguistics is the schema. The proper way of studying this is to develop a deductive system and to define its entities relationally. For example, an item of a given kind *A* is said to presuppose an item *B* in a context *C* if *A* cannot occur without *B*, but *B* can occur without *A*; for instance, an indefinite article requires a count-noun in a noun-phrase, but not vice versa. Another relation distinguished by Hjelmslev is that of interdependence, that is, of mutual presupposition, which is a relation holding, for instance, between vowels and consonants, or mood and tense. Finally, there is a relation called 'solidarity' which obtains when *A*, though compatible with *B*, can occur without it, and vice versa; for instance, this relationship holds between *accusative* and *plural* within a case system. By appeal to such relationships one can, Hjelmslev argued, characterize a linguistic-schema in purely formal terms without recourse to phonetic, phenomenological or ontological premises. Moreover, he argued that such an abstract study underlines the theory of meaning as well as that of expression-elements. For just as in phonetics 'cat' can be further analysed into /c/, /a/, /t/, so the content of 'ram' can be analysed as 'he-sheep', that of 'horse' as 'he-horse', and so on; so that in formal terms we can ask what is the set of minimal constituents and operations on them which will enable us to describe the structure of the content of the words of a language. To distinguish his conception of linguistics as the study of form, that is of a science which 'would be an algebra of language, operating with unnamed entities' (1943: 79), Hjelmselv coined the term 'glossematics'. Whether so abstract a study could be the fundamental study of a natural language, and indeed whether Saussure thought it could be, seems doubtful. But Hjelmslev's work is important because his development of the Saussurean claim that language is a form led to a formal conception of linguistics, and encouraged the comparison of the structure of natural languages with other formal structures in a number of fruitful ways.

4 Post-Bloomfieldian linguistics and distributional relations

Post-Bloomfieldian linguistics developed sophisticated methodologies for the analysis of a linguistic corpus, and, as an example, we shall concentrate on one of those developed by Z.S. Harris. This methodology did not have to be used when framing hypotheses, Harris argued, but it is needed to check the validity of claims made; and though there is no uniquely correct methodology, 'what is essential is the restriction to distribution as determining the relevance of the inquiry' (Harris 1951: 6). Since distributional relations are ones of co-occurrence or of mutual substitutability, the restriction is tantamount to the Saussurean one of the study of syntagmatic and paradigmatic relations. So that, though a corpus-based study like Harris' has no use for the conception of *la langue*, it does restrict the field of inquiry in one important respect in an analogous way to that in which Saussure did.

A distributional analysis must proceed at two levels, phonemic and morphologic. At each of these levels it must first identify the relevant units, before going on to describe the distributional relations holding between them. To

analyse *This is my house* on the morphemic level we must identify its morphemic segments. What shows, for instance, that *house* is such a segment is the fact that it is replaceable by many other *prima facie* segments, such as *car*, *hat* and *dog*, all of which can occur in other environments, such as *She has a* , *big* , and so forth. Having identified the morphemic segments in a corpus, then to give as compact a description of the corpus as possible it is necessary to group the segments into morphemic classes, and these in turn into more inclusive classes. For instance, segments occurring in the environments *The* and *s* are classed as *N*; whilst those which occur in *ed*-past are classed as *V*, within which class we could, for instance, distinguish segments occurring before *N* (*V-trans*) from those which do not (*V-intrans*). The need for more inclusive classes is shown by the fact that $N + S$ can occur whenever *N* occurs except when *N* itself is $N + S$ - we have both *the dog* and *the dogs* but not *the dogss* - so it is necessary to distinguish $N + S$ from *N* as an N^2. More inclusive classes can be defined in analogous ways, making very compact descriptions of a sentence's constituents (its 'phrase structure') possible. *The boys run*, for instance, consists of a higher order *N* followed by a *V-intrans*; whilst the higher order *N* (an *NP* in more familiar terminology) consists of a determiner and an N^2 which in turn consists of an $N + S$.

The great achievement of the post-Bloomfieldian linguistics was to propose many detailed analyses of sentences in terms of their phrase structure. This made the development of a theory of syntax possible, which is something that Saussure had notoriously failed to do. It also enabled the formulation of a strong hypothesis about *all* theories of structural linguistics, namely that because of the restriction to distributional relations, the most powerful syntactic description of a sentence available to it is of that sentence's phrase structure. Whether the kind of methodology proposed by Harris is capable of yielding such descriptions is another matter. Because of the refusal to appeal to meaning it is fatally flawed from a Saussurean perspective. And whether or not that is so, it seems that to apply his methodology Harris would often have to assume he knew something which at that stage had still to be established; it is hard to see, for instance, how on his account one could identify any morphemic segment unless one had already identified some others. But though the methodology is flawed, the crucial objection - due to Chomsky - is that the linguistic descriptions proposed are inadequate; a phrase structure analysis cannot distinguish *John is eager to please* from *John is easy to please*, for all that in the first *John* is the agent, but not in the second. So despite its many insights and achievements, structural linguistics has inherent limitations.

See also: Moscow-Tartu school; Russian literary Formalism; Structuralism; Structuralism in literary theory; Syntax

DAVID HOLDCROFT

References and further reading

Bloomfield, L. (1933) *Language*, New York: Holt.(A classic text which proposes a behaviourist approach to linguistics modelled on contemporary psychology, and sets the agenda for American structuralism.)

Caws, P. (1988) *Structuralism: The Art of the Intelligible*, Atlantic Highlands, NJ: Humanities Press.(Part IB is an excellent introduction to the linguistic basis of structuralism.)

Ducrot, O. and Todorov, T. (1972) *Dictionnaire encyclopédique des sciences du langage*, trans. C. Porter, *Encyclopaedic Dictionary of the Sciences of Language*, Oxford: Blackwell, 1981.(Contains helpful essays on functionalism, glossematics and distributionalism.)

Harris, Z.S. (1951) *Structural Linguistics*, Chicago, IL: University of Chicago Press.(A classic; Chapter 2 on methodology is an important statement.)

Hjelmslev, L. (1943) *Omkring sprogteoriens grundloeggelse*, trans. F.J. Whitfield, *Prolegomena to a Theory of Language*, Madison, WI: University of Wisconsin Press, 1961.(The best introduction to Hjelmslev's thinking.)

Joos, M. (ed.) (1958) *Reading in Linguistics: The Development of Descriptive Linguistics in America since 1925*, New York: American Council of Learned Societies.(Among other papers in this classic collection, Bloomfield (1926), Harris (1946) and Wells (1947) provide a good introduction to American structural linguistics.)

Lyons, J. (1973) 'Structuralism and Linguistics', in D. Robey (ed.) *Structuralism, An Introduction*, Oxford: Clarendon Press, 1973, 5-19.(A brilliant introductory essay.)

Vahek, J. (1966) *The Linguistic School of Prague: An Introduction to its Theory and Practice*, Bloomington, IN, and London: Indiana University Press.(An impressive introduction by a member of the School, with a detailed bibliography.)

Syntax

Syntax (more loosely, 'grammar') is the study of the properties of expressions that distinguish them as members of different linguistic categories, and 'well-formedness', that is, the ways in which expressions belonging to these categories may be combined to form larger units. Typical syntactic categories include noun, verb and sentence. Syntactic properties have played an important role not only in the study of 'natural' languages (such as English or Urdu) but also in the study of logic and computation. For example, in symbolic logic, classes of well-formed formulas are specified without mentioning what formulas (or their parts) mean, or whether they are true or false; similarly, the operations of a computer can be fruitfully specified using only syntactic properties, a fact that has a bearing on the viability of computational theories of mind.

The study of the syntax of natural language has taken on significance for philosophy in the twentieth century, partly because of the suspicion, voiced by Russell, Wittgenstein and the logical positivists, that philosophical problems often turned on misunderstandings of syntax (or the closely related notion of 'logical form'). Moreover, an idea that has been fruitfully developed since the pioneering work of Frege is that a proper understanding of syntax offers an important basis for any understanding of semantics, since the meaning of a complex expression is compositional, that is, built up from the meanings of its parts as determined by syntax.

In the mid-twentieth century, philosophical interest in the systematic study of the syntax of natural language was heightened by Noam Chomsky's work on the nature of syntactic rules and on the innateness of mental structures specific to the acquisition (or growth) of grammatical knowledge. This work formalized traditional work on grammatical categories within an approach to the theory of computability, and also revived proposals of traditional philosophical rationalists that many twentieth-century empiricists had regarded as bankrupt. Chomskian theories of grammar have become the focus of most contemporary work on syntax.

1 The need for structure

One central aim of *semantics* is to explain how the meaning of a sentence (or any other complex expression) is a function of the meanings of its parts (see Compositionality), a project that presupposes an understanding of how the parts are put together, that is, an understanding of its *syntax*.

Let us call any sequence of words a 'string'. The following strings can be put together from the list of words 'Bill', 'slept', 'soundly', 'last' and 'night':

(1) (a) Bill slept soundly last night
(b) last night Bill slept soundly
(c) *last soundly Bill slept night
(d) *night Bill soundly last slept.

There is a clear difference between, on the one hand, (1a) and (1b) and, on the other, (1c) and (1d). In ordinary talk, the former are *sentences*; the latter are not. (Asterisks indicate non-sentences.) In order for a string of words to be a sentence (in this ordinary sense) the words must be put together in a certain way. Syntax is, in part, the study of the rules (or conditions) that determine the way that sentences are structured. The order of words in a string does more than determine whether or not a string is a sentence; it also plays a role in determining meaning. Compare (2a) and (2b):

(2) (a) The dog chased the cat.
(b) The cat chased the dog.

The fact that (2a) and (2b) differ in meaning cannot be a consequence of their containing words with different meanings, since they contain precisely the same words. The difference in meaning is attributable to a difference in *word order*. In the terminology of traditional grammar, in English we understand a noun phrase preceding a verb (in the active voice) as the subject, and the noun phrase following that verb as its direct object. Not all languages mark the subject-object distinction in this way. Thus, for example Latin marks it by inflection and this allows for the possibility of much freer word order.

Initially, the following generalizations might seem reasonable: (a) sentence meaning is the product of word meaning and word order; (b) a difference in word order results in a difference in sentence meaning. But in fact neither is accurate. Although a change in word order frequently results in a change in meaning, sentences (1a) and (1b) above show that this is not always the case, so (b) is incorrect. To see that (a) is incorrect consider

(3) Mary said Bill left voluntarily.

This admits of two distinct interpretations (or readings), which can be paraphrased as (a) 'Mary said that Bill left and she said this voluntarily' and (b) 'Mary said that Bill left and that he did so voluntarily'. It is as if the word 'voluntarily' can be understood in connection with either the saying or the leaving. Sentence meaning, then, is the product of more than word meaning and word order; it is the product of word meaning and sentence structure, and there is more to the sentence structure than the linear order in which its words appear: the way in which the words are grouped is crucial. One way of cashing out this idea might be to say that 'voluntarily' can be understood as an attachment either to the sentence 'Mary said Bill left' or to the smaller sentence 'Bill left'. Using square brackets to group parts of the sentence, we can represent the two different interpretations of (3) as follows:

(4) (a) [Mary said [Bill left]] voluntarily.
(b) Mary said [[Bill left] voluntarily].

In a sense that we can make precise, the moral of an example such as (3) is that the *hierarchical* organization (grouping) of the words in a sentence is just as important as their *linear* organization. (In terms explained in §5 below, the *scope* of 'voluntarily' is different in (4a) and (4b).)

2 Syntactic categories and constituent structure

The use of syntactic theory in projects such as compositional semantics requires that the theory have some systematic way of describing the syntax of sentences. The ways in which grammatical theories can describe linguistic structure are varied, but the following exposition will illustrate the most common. Since we are not here concerned with the internal structure of words (their morphology), let us say that words are the 'ultimate constituents' of sentences. The two parts of the sentence 'Odysseus returned' correspond to the traditional distinction between subject and predicate. Of course, both subject and predicate expressions can be more complex. If we replace the subject expression 'Odysseus' in 'Odysseus returned' by 'the hero', 'no hero' or 'a hero as great as godlike Achilles', in each case the result is another sentence. Since all of these expressions are built upon nouns, they are said to belong to the category of noun phrase (NP).

If we replace the verb 'returned' in the sentence 'Odysseus returned' by 'loves Penelope', or 'shaves another customer', in each case the result is another sentence. Since expressions in predicate position are (typically) based on verbs, they are said to belong to the linguistic category of verb phrase (VP). Using 'S' for 'sentence', we can represent the syntactic structure of the sentence 'Odysseus returned' using a 'phrase structure tree' (or 'phrase marker') as follows:
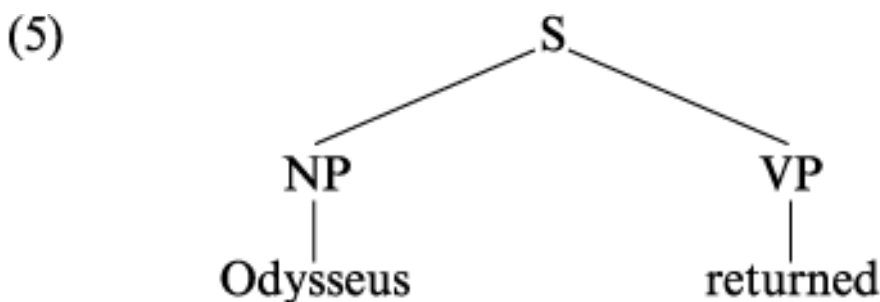
Alternatively, we can represent exactly the same syntactic information using a 'labelled bracketing':

(6) $[_S [_{NP} \text{Odysseus}][_{VP} \text{returned}]]$.

(5) and (6) are notational variants of one another, they are equivalent descriptions of the same structure - 'structural descriptions' - that is, tree notation and bracket notation are two informationally equivalent ways of specifying syntactic structure.

We can begin to characterize a set of sentences by formulating a rule to the effect that an NP followed by a VP forms a sentence, S. The standard way of doing this is to use a 'phrase structure rule' (or 'rewrite rule'):

(7) $S \Rightarrow NP + VP$.

This is read as 'S goes to NP VP'. All (7) says is that a sentence, or S, may be composed of an NP followed by a VP.

Let us turn now to the internal structures of NPs and VPs. Many NPs appear to be composed of words belonging to the traditional grammatical categories of article (for example, 'the', 'a') and noun (for example, 'cat', 'man'). Rather than using the category of article, let us use the broader category of *determiner* that includes not just 'the' and 'a', but also, for example, 'every', 'some', 'no', 'neither' and 'one'. We can represent the fact that an NP may be composed of a determiner (D) and a noun (N) using the following phrase structure rule:

(8) $NP \Rightarrow D + N$.

Now we can provide a phrase structure tree for 'The hero returned' using the rules in (7) and (8):
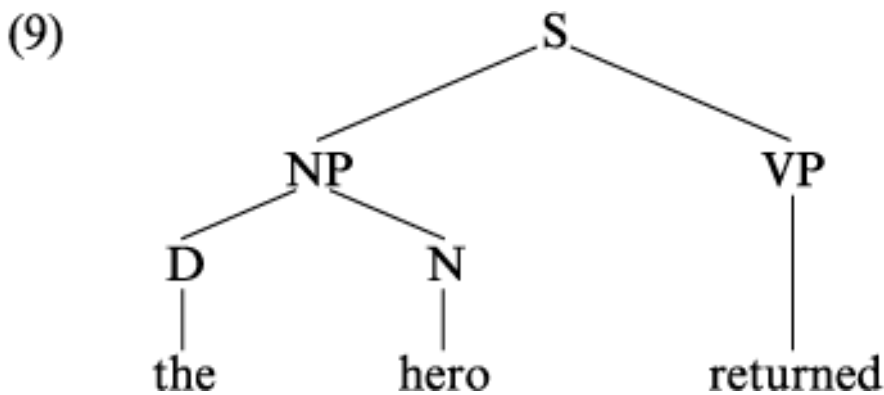
Or we can use a labelled bracketing:

(10) $[_S [_{NP} [_D \text{the}][_N \text{hero}]][_{VP} \text{returned}]]$.

Some VPs may contain NPs as parts, for example, 'loves Penelope', and 'shaves another customer' (see above). Unlike verbs such as 'return' and 'snore', verbs like 'see' and 'respect' are *transitive*: they take NPs as direct objects. We have, then, at least two types of VPs to examine: those 'headed by' (that is, built around) intransitive verbs (V_is), such as 'return' and 'snore', and those headed by transitive verbs (V_ts), such as 'respect' and 'like'. Consequently, we need at least two phrase structure rules for VPs:
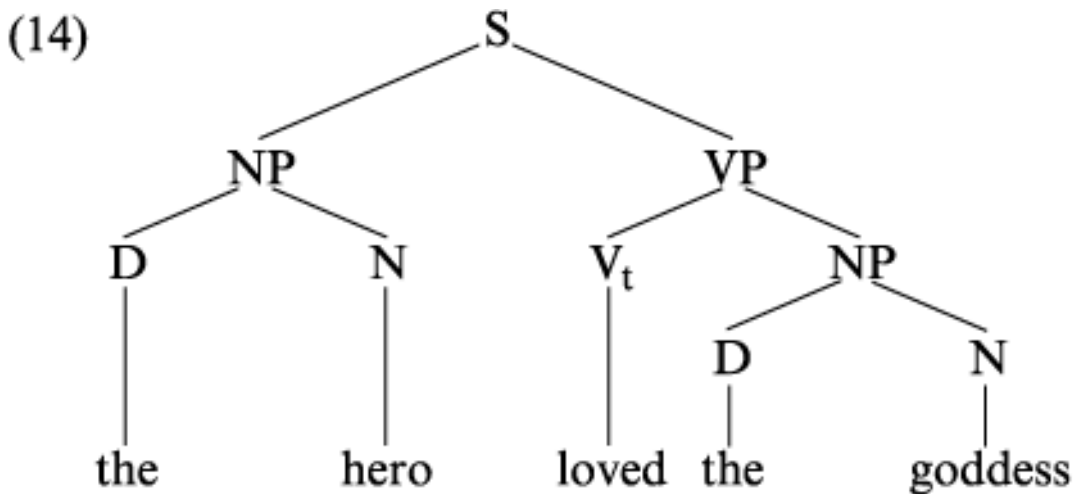
(11) $VP \Rightarrow V_i$
(12) $VP \Rightarrow V_t + NP$.

The syntax of 'The hero returned' can now be spelled out in more detail:
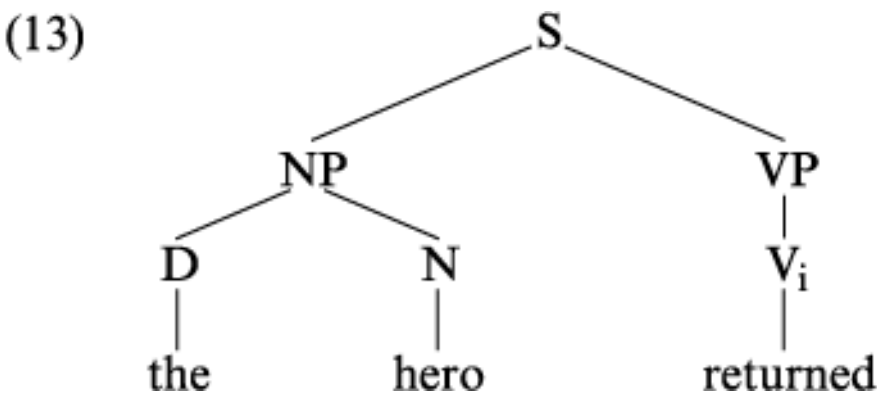
The syntax of 'The hero loved the goddess' is given by

(14)

3 Phrase structure grammars

Not only do philosophical applications of syntactic theory require a set vocabulary for describing linguistic structure, they require linguistic theories to specify well-formed linguistic structures in a finite way. Contemporary syntactic theory offers just such resources.

One aim of what Chomsky has called 'generative grammar' is to articulate a finitely statable theory that generates all and only the phrase markers of a given language. With the resources made available in the previous section, we can construct a generative grammar for a fragment of English. For purposes of illustration, we can do this by articulating a 'phrase structure grammar'. Such a grammar can be viewed as a formal system consisting of two parts: a 'lexicon' (a list of words together with a specification of the grammatical category of each word), and a set of phrase structure rules specifying how words from these categories may be put together to form sentences. In effect, then, a grammar is a theory, and like any other theory its usefulness lies in its predictive power. For any string of English words, the theory must say whether or not that string is a sentence. The following is a simple generative grammar:

Lexicon

$$PN = \{Fred, Mary, Bill\}$$
$$D = \{a, every, the\}$$
$$N = \{man, woman\}$$
$$V_i = \{left, returned\}$$
$$V_t = \{likes, respects\}$$

Phrase structure rules

$$S \Rightarrow NP + VP$$
$$NP \Rightarrow PN \text{ ('proper name')}$$
$$NP \Rightarrow D + N$$
$$VP \Rightarrow V_i$$
$$VP \Rightarrow V_t + NP$$

Notice that each of the phrase structure rules has the following form:

(15) $\alpha \Rightarrow \beta_1 + \ldots + \beta_n$.

The important feature here is that in each rule of this type exactly one symbol α appears on the left-hand side of '$\Rightarrow$'. Such rules are called 'context-free' phrase structure rules, the idea behind the terminology being that they are blind to whatever symbols may appear on either side of α, that is, they are blind to the syntactic 'context' in which α appears. The systematic use of a set of context-free rules is a way of formalizing traditional work on

grammatical categories within an elegant approach to the theory of computability (or recursive functions); a context-free phrase structure grammar is equivalent to a categorial grammar in the sense of Ajdukiewicz (1935; see Ajdukiewicz, K. §5).

In a sense to be defined, the grammar just given generates the following phrase structure tree:
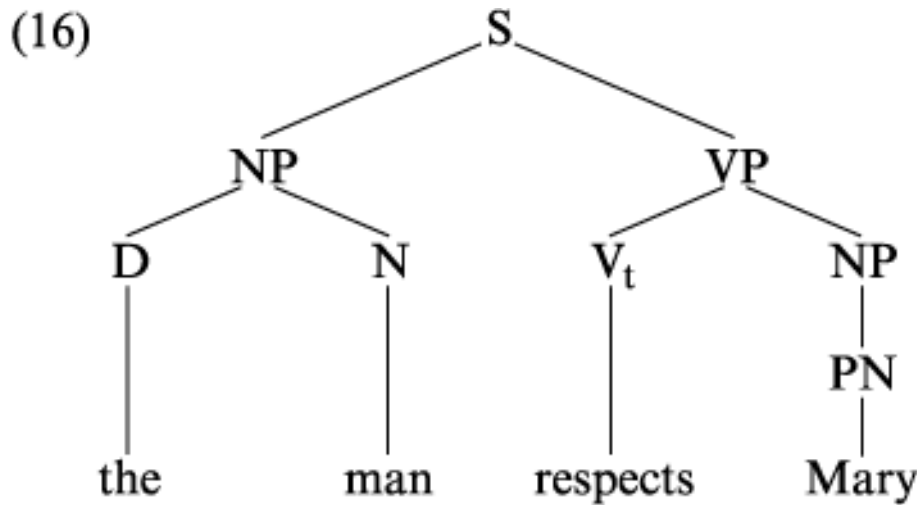
There is a specialized vocabulary in formal language theory (mathematical linguistics) for describing the parts of phrase structure trees and the relationships that obtain between the parts. We can tailor some of this vocabulary, rather informally, to suit our concerns:

(i) Each position in a tree is a *node*. A line connecting two nodes is a *branch*.

So, for example, in (1) there are twelve distinct nodes: S, NP, VP, D, N, V_t, PN, 'the', 'man', 'respects' and 'Mary'. Branches connect D and 'the', NP and D, and NP and N.

(ii) The S node at the 'top' of a tree is the *root node* of that tree. The words at the 'bottom' of a tree are its *leaf nodes* (or *terminal nodes*).

We are concerned with (a fragment of) English, so every leaf node will be a word of English.

(iii) A sequence of branches that connects two nodes is a *path*. Node β is *properly dominated* by node α if and only if the path from β back to the root node passes through α. The set of nodes properly dominated by node α is the *proper domain* of α.

Example: in (16) there is a path between D and 'the', and a path between S and D. Indeed, there is a path between any two distinct nodes. Additionally, D properly dominates 'the', NP properly dominates D, N, 'the' and 'man', and S properly dominates every node in the tree (apart from itself). (It dominates but does not properly dominate itself, just as a set is a subset but not a proper subset of itself.)

(iv) Each node in a tree corresponds to a *constituent* of the sentence in question.

So, for example, in (16) the NP node corresponds to the constituent 'the man'; and the VP node corresponds to the constituent 'respects Mary'. So whereas 'the', 'the man', 'respects' and 'respects Mary' are constituents of (16), 'man respects', 'man respects Mary' and 'the man respects' are not.

(v) If Γ is a phrase structure grammar and τ is a tree, then Γ *generates* τ if and only if
(a) the root node of τ is S;
(b) every terminal node of τ is in the lexicon for Γ; and
(c) every step from the root node to the string of terminal nodes is licensed by one of the phrase structure rules of Γ.
(vi) A *string* of Γ is a linear sequence of words taken from the lexicon for Γ. Γ *generates a string* Σ if and only if Γ generates a tree for Σ.

(vii) The set of strings generated by a grammar Γ is the *language* L_Γ of Γ, that is, the set of *sentences* of L_Γ.

An interesting feature of (vii) is that it appropriates the English words 'language' and 'sentence' and assigns them technical meanings. Later, the theoretical use of these words will be specified precisely to satisfy theoretical demands that arise when syntax and semantics come together.

The picture just presented is deliberately simplified. Nevertheless, it is widely held that context-free grammars of this form are inadequate to describe natural languages. In addition to context-free rules of the type we have just discussed, Chomsky (1957, 1965) posited 'transformational' rules that operated on the trees generated by context-free phrase structure grammars to produce trees that could not be produced in any satisfactory way by phrase structure rules alone. For example, Chomsky wanted to explain the syntactic and semantic relationships between the active and passive voice and accomplished this by postulating a transformational rule that derived a phrase structure tree for a passive sentence from the tree for its corresponding active. Much work by generative linguists in the 1970s sought to impose a rigorous set of constraints on possible transformational rules. Beginning with Chomsky (1981), generative linguists began to explore alternative ways of generating sentences using methods that depart from those of traditional phrase structure grammars.

4 Recursion and sentential verbs

The grammar above was finite, as was the language that it generated (it consists of a finite number of sentences). There are, however, finite grammars that generate infinite languages. This can be seen easily enough in connection with philosophically interesting verbs such as 'believe', 'know', 'prove', 'remember' and 'say', which can be used in ways that preclude classifying them as either V_is or V_ts. Consider the following sentences:

(17) Mary believes (that) Fred left.
(18) Mary said (that) Fred left voluntarily.

The VPs in these sentences seem to consist of a verb combined with a whole sentence. Since they seem to take whole sentences as their complements, these can be called 'sentential verbs' (V_s). Let us now add some V_ss to our lexicon:

$V_s = \{$believe, know, doubt, say, suggest, realize, remember, prove$\}$.

(Some V_ss are also known as '*psychological* verbs' or 'verbs of propositional attitude', a label due to Russell - see Propositional attitude statements; §5 below.) We can now add the following phrase structure rule to produce a new grammar:

(19) $VP \Rightarrow V_s + S$.

The resulting grammar generates the following phrase marker:
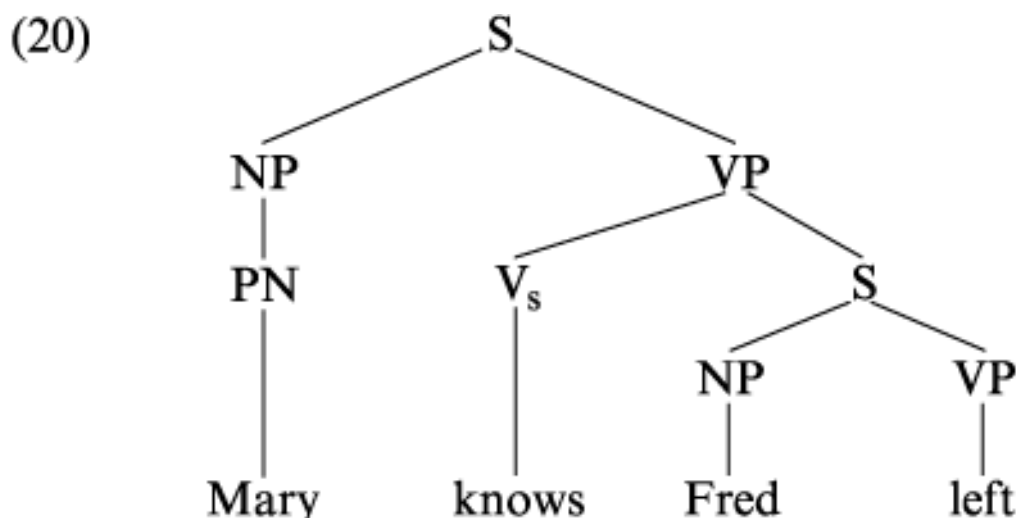
As this tree diagram reveals, the rules ' $S \Rightarrow NP + VP$' and ' $VP \Rightarrow V_s + S$' taken together give our grammar the capacity to generate trees in which S nodes are properly dominated by other S nodes. More precisely, our grammar generates trees in which an S node properly dominates a VP node that properly dominates another S node that properly dominates another VP node that…. So, for example, the grammar generates trees for both of the following:

(21) Bill knows Mary thinks Fred left.
(22) Mary doubts Fred knows Bill said Mary left.

Clearly, there is no end to the list of sentences generated: if our grammar generates a sentence ϕ, it also generates every sentence of the form

$[_S [_{NP} \ldots][_{VP} [_{V_s} \ldots]\phi]]$.

Despite containing a finite set of rules, it generates an infinite set of sentences (the language generated is infinite). A structure such as (20) in which a node of category X properly dominates a distinct node also of category X is a *recursive* structure. We now say that our grammar *recursively generates* an infinite set of sentences. The recursive nature of natural language is further revealed once connectives such as 'and', 'or' and 'but' are introduced. A primary function of connectives is to join two sentences - they may also be used to connect, for example, NPs and VPs - in such a way that the result is a larger sentence, as in

(23) Bill left and Mary returned.
(24) Bill snores or Mary snores.

The following phrase structure rule can therefore be postulated:

(25) $S \Rightarrow S \ CONN \ S$,

where CONN is the category 'connective'. This rule exploits the possibility of having three symbols on the right-hand side of ' $\Rightarrow$', but it is still a context-free phrase structure rule of the form given in (15). Recursive structures are also found in (for example) NPs containing relative clauses such as 'the man who loves the woman who found the child who…'.

5 Scope

There are a number of semantic issues of independent philosophical importance that can be illuminated by appealing to syntactic structure. The nature of scope in natural language is an instructive case study (see Scope). The following strings are all ambiguous:

(26) (a) everyone strives for some good
(b) if I know that *p* then necessarily *p*
(c) the number of planets is necessarily greater than seven
(d) Bill returned or Fred returned and Mary left
(e) George thought the King wasn't the King
(f) Mary said Bill left voluntarily.

For example, someone uttering (26a) might be understood as saying that there is some particular good such that everyone strives for it or merely that everyone strives for some good or other. All of these examples involve ambiguities of scope that can be explicated within sophisticated syntactic theories. For present concerns, we restrict ourselves to a simple case mentioned earlier involving the scope of the adverb 'voluntarily' as it occurs in a string such as (26f), which is ambiguous between readings corresponding to the syntactic structures given earlier by (4a) and (4b). To provide a general characterization of scope in terms of phrase structure is straightforward once we introduce some new notions. First,

(viii) Node β is *immediately dominated* by node α if and only if β is properly dominated by α and there is no intervening node γ, that is, no node γ such that α properly dominates γ and γ properly dominates β.

In the following diagram, A properly dominates every other node; in addition it immediately dominates B and C,

but not D, E, F or G:
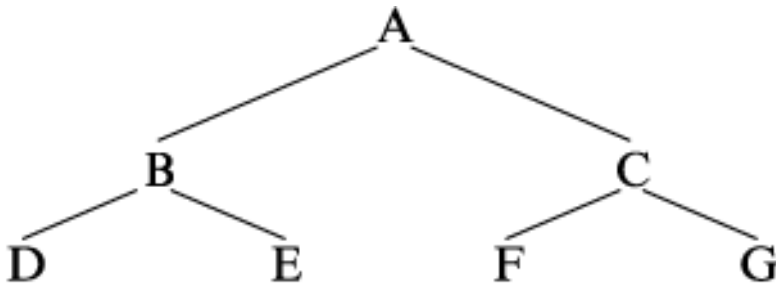
(ix) If distinct nodes β and γ are immediately dominated by node α , then α is a *branching node*.

A, B and C are branching nodes; D, E, F and G are not. We can now define 'scope' in a phrase structure tree:

(x) The *scope* of node α is (the constituent corresponding to) the first branching node properly dominating α . Node β is *within the scope* of node α if and only if the first branching node properly dominating α properly dominates β .

So in (20), the scope of 'Mary' is the entire sentence, whereas the scope of 'Fred' is the smaller sentence 'Fred left'.

(xi) If node β is within the scope of node α and node α is not within the scope of node β , then α has larger scope than β (and β has smaller scope than α).

Thus 'Mary' has larger scope than 'Fred' in (20). The scope of 'knows' in this phrase marker is the VP 'knows Fred left'. Thus the name 'Fred' occurs within the scope of 'knows', but the name 'Mary' does not. This is widely held to have important logical consequences. If Mary is Elizabeth - that is, if 'Mary' and 'Elizabeth' are two names of the same person - then the truth of (21) guarantees the truth of

(27) Elizabeth thinks Fred left.

Now suppose that Fred is Bert. Does the truth of (21) guarantee the truth of (28)?

(28) Mary thinks Bert left.

Following Frege, many philosophers believe it does not. (Mary may not know that Fred is Bert.) If this is correct, then we can describe the situation syntactically: the fact that 'Fred' appears within the scope of a 'psychological' verb (in this case, 'think') means there is no guarantee that 'Fred' can be replaced by a coreferring name to produce a sentence with the same truth-value as the original. A name occupying such a position is said to occur in a context that is 'referentially opaque' (see Propositional attitude statements §§1-2).

6 Ambiguities of scope

We turn now to an example involving what is usually called a 'structural ambiguity' or an 'ambiguity of scope'. The latter label (due to Russell) is more dominant in philosophy, logic, and mathematics. As Russell pointed out, much bad philosophy has resulted from inattention to scope ambiguities. Indeed, uncovering and avoiding them are skills all philosophers must acquire. Davidson has emphasized the importance of an understanding of the semantics of adverbs to a number of philosophical questions involving actions and events. Earlier, the fact that the following string is ambiguous was used to call attention to the grouping of words within a sentence:

(3) Mary said Bill left voluntarily.

It was suggested that the ambiguity in question was structural in that the words can legitimately be grouped in two different ways:

(4) (a) [Mary said [Bill left]] voluntarily.
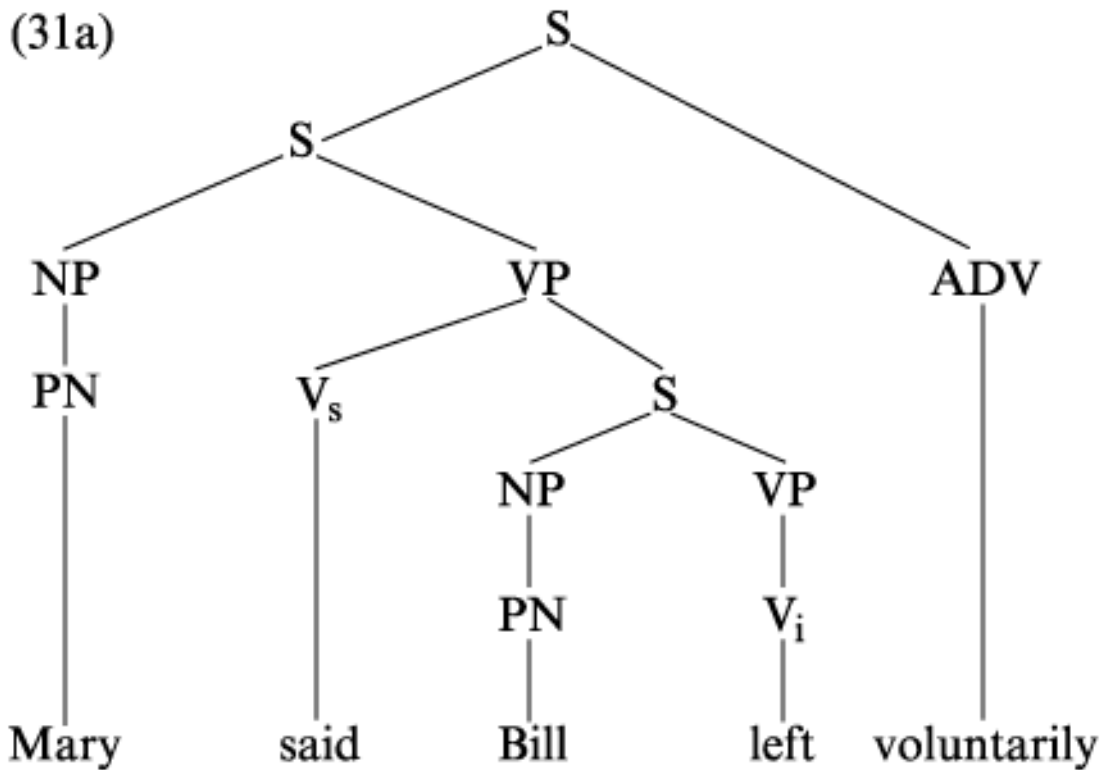(b) Mary said [[Bill left] voluntarily].

We can implement this suggestion by providing distinct phrase markers corresponding to (4a) and (4b). We will consider two rival ways of doing this. The word 'voluntarily' belongs to the category of adverb, so let us add to our lexicon the following:

(29) ADV = {voluntarily, hastily, discreetly}.

In line with the informal account of the ambiguity suggested earlier, the first attempt to characterize the syntax of adverbs involves positing the following phrase structure rule:

(30) $S \Rightarrow S + ADV$.

The resulting grammar generates the following distinct phrase markers for (3):

(31a)

```
                                    S
                        S                        ADV
              NP              VP
              PN         V_s           S
                                   NP       VP
                                   PN       V_i
          Mary        said      Bill      left   voluntarily
```

(31b)
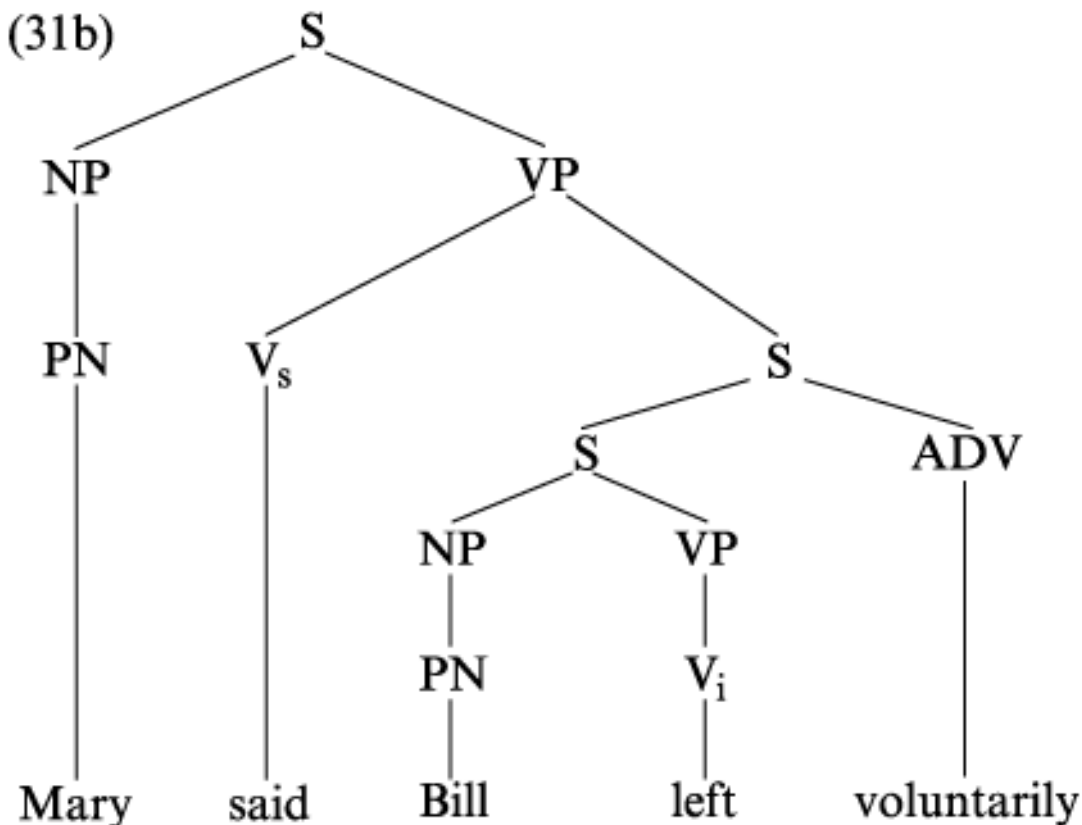
In (31a) the scope - as defined above - of 'voluntarily' is the entire sentence. Thus the verb 'said' is within its scope. By contrast, in (31b) the scope of 'voluntarily' is just the smaller sentence 'Bill left voluntarily'. Thus the verb 'said' is not within its scope, but the verb 'left' is. On the assumption that, as the etymology suggests, an adverb modifies a verb in some way, the verb it modifies is the main verb of the sentence to which it is attached. Thus (31a) corresponds to the reading of (3) that might be loosely paraphrased as 'Of her own volition, Mary said that Bill left' and (31b) corresponds to the reading that might be paraphrased as 'Mary said that Bill left as a matter of his own volition'. (Of course, until an account is provided of how syntactic structure contributes to meaning, the pairing of distinct phrase markers with distinct readings of a string is largely intuitive.)

We now make two observations about our definition of scope. First, by the definition provided, every constituent has a scope; but there are only certain types of constituents whose scopes will interest us (usually, connectives, quantifiers, verbs and adverbs). Second, notice that it is not always the case that for two constituents one is within the scope of the other.

It is worth asking whether this is the best way of characterizing the syntax of sentences containing adverbs. An alternative account emerges if we take the etymology of 'adverb' more seriously and view ADVs as attaching to VP nodes rather than S nodes. Only detailed empirical and logical work will tell which of these analyses is better. The best theory may tell us that there are two distinct types of ADV: one sort that attaches to an S node, another that attaches to VP. There is no need for us to settle this difficult matter here. Suffice to say, that if we find that claims about the semantic properties of adverbs are doing some philosophical work, we will have to look carefully at their syntax.

More than a few philosophical muddles have been due to confusions over the scope of operators in natural language sentences. Utilizing the resources of syntactic theory, it becomes possible for us to study scope relations in a more rigorous fashion, and to marshal arguments for and against certain proposals about the logical form of natural language. Indeed, whenever we encounter philosophical claims which rely upon the semantic properties of quantifiers, adverbs and so on, we will want to look carefully at their syntactic properties. This specific point is just an instance of a quite general principle concerning the relationship between the philosophy of language and theoretical syntax: the former cannot properly proceed independently of the latter.

See also: Analytical philosophy; Chomsky, N.; Language of thought; Logical form; Mind, computational theories

of; Scope

STEPHEN NEALE

References and further reading

Ajdukiewicz, K. (1935) 'Die Syntaktische Konnexität', *Studia Philosophica* 1: 1-27; trans. 'Syntactic Connexion', in S. McCall (ed.) *Polish Logic*, Oxford: Clarendon Press, 1967.(A full statement of the theory of categorial grammar.)

Chomsky, N. (1957) *Syntactic Structures*, The Hague: Mouton.(Early statement of the formal limitations of various approaches to syntax, mathematical properties of different types of grammars, and the theory of transformational grammar. Very readable.)

Chomsky, N. (1965) *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press.(Detailed statement of Chomskian methodology, and Chomsky's 'Standard Theory'.)

Chomsky, N. (1981) *Lectures on Government and Binding*, Dordrecht: Foris.(Detailed statement of major components of Chomsky's 'principles and parameters' approach to syntax that moves well beyond the transformational approach of his earlier work. Hard going.)

Chomsky, N. (1986) *Knowledge of Language*, New York: Praeger.(Very readable account of the philosophical and technical details of the 'principles and parameters' approach.)

Davidson, D. (1980) 'The Logical Form of Action Sentences', in *Essays on Actions and Events*, Oxford: Clarendon Press.(A proposal about the logical form of sentences involving action verbs.)

Lewis, D.K. (1983) 'General Semantics', in *Philosophical Papers*, vol. 1, Oxford: Oxford University Press. (Accessible exposition of a categorial grammar for a part of English and a corresponding compositional semantics.)

May, R. (1985) *Logical Form: Its Structure and Derivation*, Cambridge, MA: MIT Press.(First sustained attempt to connect work in Chomsky (1981) with traditional philosophical work on 'logical form'.)

Neale, S. (1994) 'Logical Form and LF', in C. Otero (ed.) *Noam Chomsky: Critical Assessments*, London: Routledge & Kegan Paul.(An attempt to connect Chomsky's syntactic programme to work on meaning, reference and logical form.)

Russell, B.A.W. (1905) 'On Denoting', *Mind* 14: 479-93; repr. in *Logic and Knowledge: Essays 1901-1950*, ed. R.C. Marsh, London: Allen & Unwin, 1956, 41-56.(The classic theory of definite descriptions that provided a paradigm for analysis of logical form.)

Sells, P. (1985) *Lectures on Contemporary Syntactic Theories*, Stanford, CA: Center for the Study of Language and Information, Stanford University.(Accessible overview and exposition of competing approaches to generative grammar.)

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, trans. D. Pears and B. McGuinness, London: Routledge & Kegan Paul, 1961.(An early, philosophically ambitious theory of logical form.)

Type/token distinction

The type/token distinction is related to that between universals and particulars. C.S. Peirce introduced the terms 'type' and 'token', and illustrated the distinction by pointing to two senses of 'word': in one, there is only one word 'the' in the English language; in the other, there are numerous words 'the' on the physical page you are now looking at. The latter are spatiotemporal objects composed of ink; they are said to be word tokens of the former, which is said to be the word type and is abstract. Phonemes, letters and sentences also come in types and tokens.

The distinction between 'type' and 'token', or something very like it, seems to be applicable beyond language to, for example, Beethoven's *Fifth Symphony* and performances of it, the grizzly bear and specimens of it, the Kentucky Derby and runnings of it, and the bubonic plague and outbreaks of it. The type/token distinction is important to linguistics, logic, aesthetics and philosophy of science. In philosophy of mind it is critical to distinguish types of events/states/processes from tokens of them, because those who identify mental events with physical events agree that every mental event token is a physical event token, but divide over whether mental event types, for example, pain, are identical to physical event types, for example, C-fibre stimulation (see Mind, identity theory of). Just what types are, whether they exist, whether they are 'present in' their tokens and whether all types have tokens are matters of controversy.

To see that we often refer to types, consider the grizzly bear. At one time its US range was most of the west and it numbered 10,000 in California alone. At the end of the twentieth century its range is Montana, Wyoming and Idaho, and it numbers fewer than 1,000. But no particular flesh and blood bear numbers 1,000 or once had a range comprising most of the western USA; if anything, it is a type which does. Similarly, Old Glory had twenty-nine stars in 1846 but fifty in 1996, whether or not any particular American flag underwent such a transformation; Old Glory is also a type.

This last example points to a related notion, that of an 'occurrence'. The stars on Old Glory number fifty; but fifty star *types* or fifty star *tokens*? In fact it cannot be either: not star types, because all the stars on the flag are of the same (five-pointed) type; nor star tokens, because tokens are concrete and the flag in question is abstract and cannot contain concrete parts. Old Glory contains fifty occurrences of the (five-pointed) star. Similarly, the letter 'x' (the very same letter type) occurs three times in the formula (type) '$(\forall x)(Fx \rightarrow Gx)$'. Thus the notion of an occurrence of 'x' must not be confused with the notion of a token of 'x' (although it often is so confused). The notion of 'an occurrence of x in y' involves not only x and y, but also how x is situated in y. If we think of a formula as a sequence, then the air of mystery over how the same identical thing can occur twice vanishes. Even concrete particulars can occur more than once (for example, the same person occurs twice in the sequence of New Jersey million dollar lottery winners).

Tokens are particulars. Are types universals? They are if having instances makes something a universal. But in many ordinary and theoretical contexts, terms which refer to types function not as predicates but as singular terms, in sentences that permit existential generalization. ('The grizzly bear ranged over most of the west' entails 'Something ranged over most of the west', for example.) As values of bound first-order variables, they meet a Quinean necessary condition for existence, and thus may also be viewed as particulars (see Ontological commitment §1).

Do types exist? If types are universals, then the debate over whether they exist goes back at least to Plato. Still, as the preceding paragraph noted, types seem to have an existential advantage over more traditional universals. (Quine held that expression types exist, though not more traditional universals (1987).) Nominalists who deny that there are any abstract objects argue that this is an illusion; that talk about types is just shorthand for talk about tokens. The matter cannot be resolved here, but it may be useful to examine one popular nominalistic account, namely, that every reference to a type (for example, 'The grizzly bear is ferocious') can be replaced without change of information by quantification over all tokens (for example, 'Every grizzly bear is ferocious'). One problem for this account is the existence of numerous counterexamples. 'The grizzly bear ranged over most of the west', for example, cannot be analysed as 'Every grizzly bear ranged over most of the west'. (Perhaps it can be analysed in some other 'type-free' way, but the point is that the nominalistic account under consideration cannot be the systematic one needed to eliminate all the many references to types which occur in our ordinary and

scientific talk.)

A second problem is created by the implication that all types have tokens. Peirce (1931-58) claimed they do, but many thinkers since have denied this. Chomsky (1957), for example, claimed that there are sentences that never have been or will be instantiated - infinitely many. A third problem arises when we try to find a replacement for even a simple true sentence about, say, the word 'one', given the myriad senses it has, and the myriad forms its tokens may take: printed in ink in any number of fonts and handwritings, raised in Braille, incised in marble, existing briefly as so many pixels of light on a computer screen, electronic strings of dots and dashes, smoke signals, hand signals, individual air disturbances produced by human vocal cords, electromagnetic pulses on phone lines, pronounced (and mispronounced) in countless ways in countless accents. Even an (appropriately surrounded) empty space can be a token of 'one'. About the only thing all and only tokens of 'one' clearly have in common is being tokens of 'one' - the type, that is; if that word (type) does not exist, then they would seem to have nothing in common and hence no appropriate quantification could even be formulated.

See also: Abstract objects; Logical and mathematical terms, glossary ofUniversals

LINDA WETZEL

References and further reading

Block, N. (ed.) (1980) *Readings in Philosophy of Psychology*, Cambridge, MA: Harvard University Press, vol.1. (Includes numerous important articles in philosophy of mind pertinent to the debate between token materialists and type materialists.)

Bromberger, S. (1992) 'Types and Tokens in Linguistics', in *On What We Know We Don't Know*, Chicago, IL: University of Chicago Press.(A fine attempt to explain the relationship between types and tokens, in order to explain how linguistics is an empirical science although about abstract types.)

Chomsky, N. (1957) *Syntactic Structures*, The Hague: Mouton.(The classic attack on structuralism which started the Chomskian revolution in linguistics.)

Goodman, N. (1951) *The Structure of Appearance*, Dordrecht: Reidel, 3rd edn, 1977. (Although not a book about types, Goodman's attempt to do without them produces many instructive passages.)

Goodman, N. and Quine, W.V. (1947) 'Steps Toward a Constructive Nominalism', *Journal of Symbolic Logic* 12: 105-22; repr. in *Problems and Projects*, Indianapolis, IN: Bobbs-Merrill, 1972.(A nominalist attempt at proof theory without expression types.)

Hale, R.J. (1987) *Abstract Objects*, Oxford and New York: Blackwell.(An excellent discussion about abstract objects generally, including types, and our reasons for being justified in concluding they exist.)

Hutton, C. (1990) *Abstraction and Instance: The Type-Token Relation in Linguistic Theory*, Oxford: Pergamon. (By means of an impressive survey of the relevant literature, Hutton argues that linguistic theories rely uncritically on the type/token distinction (and on related notions such as that of *langue-parole* and competence-performance) and that all such accounts are problematic.)

Katz, J.J. (1981) *Language and Other Abstract Objects*, Totowa, NJ: Rowman & Littlefield.(A robust defence of languages and sentences as abstract objects.)

Peirce, C.S. (1931-58) *Collected Papers of Charles Sanders Peirce*, ed. C. Hartshorne and P. Weiss, Cambridge, MA: Harvard University Press, vol. 4, para. 537; vol. 8, para. 334; vol. 2, para. 243-6.(The eight-volume collection of Peirce's papers.)

Quine, W.V. (1987) *Quiddities: An Intermittently Philosophical Dictionary*, Cambridge, MA: Harvard University Press, 216-19.(This is Quine distilled down to eighty-three alphabetically arranged entries, including a two-page entry on 'Type versus Token'.)

Wetzel, L. (1993) 'What Are Occurrences of Expressions?', *Journal of Philosophical Logic* 22: 215-20.(Provides an answer to the title question.)

Wollheim, R. (1968) *Art and Its Objects*, New York: Harper & Row.(Argues that many works of art cannot be physical objects/events or classes thereof, but are types that none the less are 'present in' their tokens.)

Wolterstorff, N. (1975) 'Toward an Ontology of Art Works', *Noûs* 9: 115-42. (Wolterstorff argues that many works of art are types, or 'kinds', and proposes a suitable analysis of the notion of a kind.)

Universal language

Most often associated with attempts to establish an international language such as Esperanto, the idea of a universal language is rooted in the biblical claim of an original language common to all human beings. The idea received its most thorough investigation during the seventeenth century. Drawing on the example of Chinese characters, early schemes involved a system of written signs that would allow communication between speakers of different languages. Later thinkers argued for the importance of an ideal 'philosophical language' in which the structure of signs exactly mirrored the structure of reality. While such projects fell short of their authors' expectations, their influence can be discerned in the formalisms of modern logic and science.

1 Adamic language

The earliest source for the idea of a universal language is the biblical story of an original language used by Adam to name the different species of animals created by God (*Genesis* 2: 19-20). In both the Jewish and Christian traditions, it was commonly assumed that this 'Adamic' language expressed Adam's perfect knowledge of things prior to the Fall: each name of a creature exactly conveyed its essence. It was natural to suppose a corruption of this language at the Fall; however, the crucial event in the loss of a universal language is told in the story of Babel, when 'the Lord confused the language of all the earth' (*Genesis* 11: 9). Thereafter, human beings were condemned to speak a multitude of languages and to suffer the pain of mutual incomprehension.

This biblical background remained central to conceptions of a universal language well into the seventeenth century. Numerous attempts were made to identify and recover the original language of Adam. Among Jewish and Christian thinkers alike, Hebrew retained the greatest claim to this honour, although other candidates were proposed (including Latin, Chinese, Dutch and Swedish). The attempt to recover the Adamic language, and the knowledge implicit in it, was significantly influenced by the Jewish kabbalah, which assigned a mystical significance to the letters of the Hebrew alphabet and prescribed techniques for their interpretation and manipulation (see Kabbalah). Another important source was the thirteenth-century theologian Ramon Llull, who stressed a method of achieving universal knowledge through combinations of letters signifying fundamental categories of reality (see Llull, R. §2).

2 Artificial schemes

During the seventeenth century, there occurred an explosion of interest in universal language schemes, conceived for the first time as systems of artificial signs or 'characters' constructed by human beings as a means of overcoming the limitations of natural languages. The causes of this newfound fascination with universal languages are complex. Clearly, a role must be assigned to the growing importance of vernacular languages and the decline of Latin as a shared medium of commerce, scholarship and diplomacy. An increase in millenarian religious sentiment, particularly in England, renewed the call for a recovery of the Adamic language. Finally, and most importantly, there was a growing recognition of the significance of language as a factor in the acquisition of scientific knowledge.

One of the earliest and most influential statements on the topic was Francis Bacon's reference in *The Advancement of Learning* (1605 II: ch. 16) to languages such as Chinese, whose users 'write in characters real, which express neither letters nor words in gross, but things or notions'. To many seventeenth-century thinkers, Chinese provided a model of what a universal language might be: a single set of characters that could be pronounced differently in different languages, but which when written would offer a shared basis for understanding. Implicit in Bacon's remark, however, was a further idea that pointed toward the scientific value of such a language. In his *New Organon* (1620 I: aph. 59), Bacon complained of how words commonly obscure 'the true divisions of nature'. It was a short step from this to the idea that these 'true divisions' might be better represented in a language composed of 'real characters', which directly expressed 'things or notions' (see Bacon, F. §§4-5).

Most early proponents of artificial language schemes stressed the practical value of their inventions as instruments of communication. Their works were primarily attempts to devise a system of writing, modelled variously on Chinese characters, cryptographic codes or shorthand notation, whereby synonymous words in different languages would be represented by a common sign. It was not long, however, before the more ambitious idea of a

'philosophical language' took hold. In a 1629 letter to Marin Mersenne, Descartes had already expressed scepticism concerning the usefulness of artificial languages of the first sort. However, he went on to postulate another kind of language in which ideas would be represented so clearly that errors of judgment would be 'almost impossible'. To realize such a language, all of our thoughts would first have to be given a proper order 'like the natural order of the numbers'; and this presupposes the 'true philosophy', by which the analysis and ordering of thoughts would be carried out. Although Descartes pursues the plan no further, he is optimistic that 'such a language is possible and that the knowledge on which it depends can be discovered'.

A philosophical language of the sort envisioned by Descartes is described in Mersenne's *Harmonie Universelle* (1636) and was pursued in many later works, culminating in George Dalgarno's *Ars Signorum* (1661) and John Wilkins' *An Essay Towards a Real Character and a Philosophical Language* (1668). Such schemes typically consisted of two parts: a system of categories summarizing the 'true divisions' of nature, and a set of characters suitable for representing these categories and the elements within them. Although informed by the discoveries of seventeenth-century science, the first part of the scheme was strongly indebted to the systems of categories propounded by Aristotle and medieval philosophers. In his *Essay*, Wilkins begins with forty genera (classified as transcendentals, substances, quantities, qualities, actions and relations), each of which he subdivides into its 'proper differences and species'. He then proposes two ways of representing the composition of concepts from their respective genus, difference and species: (1) a real character, or system of ideographic signs formed from combinations of vertical and horizontal lines; (2) a speakable philosophical language, consisting of novel combinations of syllables, consonants and vowels. In both cases, additional signs must be added to play the role of particles underwriting the grammatical structure of the language.

3 Leibniz

The idea of a universal language received its final significant development in the seventeenth century in the writings of Gottfried Wilhelm Leibniz. Although indebted to the efforts of Dalgarno and Wilkins, Leibniz criticizes these authors for their reliance on what he sees as an arbitrary number of basic categories. Harking back to Descartes, Leibniz locates the key to a philosophical language in an analysis by which every concept would be broken down into its simplest elements, the 'alphabet of human thoughts'. With these simples identified, and appropriate characters assigned to each, we would, in effect, possess the Adamic language: one in which the structure of every character perfectly mirrors the structure of reality.

Leibniz's plans for a universal language are among the most ambitious devised; in practice, however, they met with limited success. Early on, he recognized that an analysis of concepts into ultimate simples lies beyond the power of the human mind. What we are left with, then, are classificatory systems, like that of Wilkins, which represent our best attempts to order reality as we conceive it. Leibniz's greatest achievement in the area lies in his *specieuse générale* or 'general science of forms', a collection of calculi designed to support the formalization of reasoning in every branch of knowledge. Anticipating twentieth-century developments in logic and methodology, Leibniz's studies represent one of the enduring legacies of the universal language movement (see Leibniz, G.W. §10).

See also: Formal languages and systems

DONALD RUTHERFORD

References and further reading

Bacon, F. (1605) *The Advancement of Learning*, ed. G.W. Kitchin, London: Dent, 1976.(Bacon's discussion of 'characters real' occurs in book 2, chapter 16.)

Bacon, F. (1620) *The New Organon*, ed. F.H. Anderson, New York: Macmillan, 1960. (Bacon criticizes words as 'idols of the marketplace' in book I, aphorisms 43, 59-60.)

Dalgarno, G. (1834) *The Works of George Dalgarno*, Edinburgh; repr. New York: AMS Press, 1971.(Contains all of Dalgarno's writings on artificial language schemes, including his 1661 *Ars Signorum*.)

Descartes, R. (1991) *The Philosophical Writings of Descartes*, vol. 3 (The Correspondence), trans. J. Cottingham, R. Stoothof, D. Murdoch and A. Kenny, Cambridge: Cambridge University Press.(Descartes' letter to Mersenne is dated 20 November 1629.)

Katz, D.S. (1982) *Philo-Semitism and the Readmission of the Jews to England (1603-1655)*, Oxford: Clarendon Press.(Chapter 2 argues persuasively for the importance of the idea of an Adamic language in

seventeenth-century England.)

Knowlson, J. (1975) *Universal Language Schemes in England and France 1600-1800*, Toronto, Ont.: University of Toronto Press.(The best survey of the topic; contains a valuable checklist of universal language schemes in the seventeenth and eighteenth centuries.)

Leibniz, G.W. (1966) *Logical Papers*, trans. and ed. G.H.R. Parkinson, Oxford: Clarendon Press.(The fullest collection in English of Leibniz's writings on logic and language; includes selections from his 1666 essay *On the Art of Combinations*, in which appears his earliest plan for a universal language.)

Mersenne, M. (1636) *Harmonie Universelle*, Paris; repr. Paris: Centre National de la Recherche Scientifique, 1986.(Book I, proposition 24 contains Mersenne's proposal for a 'perfect language', in which mathematical and scientific truths would be expressed by combinations of musical tones.)

Rutherford, D. (1995) 'Philosophy and Language', in N. Jolley (ed.) *The Cambridge Companion to Leibniz*, New York: Cambridge University Press.(A comprehensive review of Leibniz's interests in language, including the idea of a universal language.)

Salmon, V. (1988) *The Study of Language in 17th-Century England*, Amsterdam: John Benjamins.(A collection of ground-breaking essays on the projects of Dalgarno, Wilkins and their contemporaries; includes a bibliography of recent work in the area.)

Slaughter, M.M. (1982) *Universal Languages and Scientific Taxonomy in the Seventeenth Century*, Cambridge: Cambridge University Press.(An important study of the relationship between seventeenth-century scientific thinking and the idea of a universal language.)

Wilkins, J. (1668) *An Essay towards a Real Character, and a Philosophical Language*, London; repr. in the series *English Linguistics 1500-1800*, number 119, Menston: Scolar Press, 1968.(Wilkins' attempt at constructing a universal language, discussed in §2 above.)

Use/mention distinction and quotation

Speakers 'use' the expressions they utter and 'mention' the individuals they talk about. Connected with the roles of used expressions and mentioned individuals is a way of uniting them and a characteristic mistake involving them. Usually the expression used in an utterance will not be the same as the individual mentioned, but the two can be made to converge. The means is quotation. Quotation is a special usage in which an expression is used to mention itself. A failure to distinguish between the roles of used expressions and mentioned individuals can lead to mistakes. Such mistakes are called use/mention confusions. In themselves use/mention confusions are a minor linguistic faux pas*, but under unfavourable conditions, they have the potential to cause greater problems.*

1 Use and mention

Though philosophers since Plato have distinguished between language and the world it is used to describe, the terms 'use' and 'mention' were first applied to these categories by the philosopher and logician Willard Quine in 1940. Quine's choice of the terms was intended to encourage associations with their ordinary meanings. To *use* an expression linguistically is to use it as a tool; to employ it towards a linguistic end. The ends are performances of speech acts, such as making claims, asking questions and giving responses. In contrast, to *mention* an individual is to say something about it; to make it the topic of conversation between speaker and listener. An individual mentioned in a speech act is one whose characteristics help determine how the act is to be evaluated: whether a claim is true, a question pertinent or a response correct depends on the individuals mentioned in the act and what is said about them.

In every communicative setting, some expression is used and, in most, some individual is mentioned. For example, a speaker uttering (1) below succeeds in making a claim about Tucson, the capital city of the state of Arizona:

(1) Tucson is dry.

By using the words 'Tucson', 'is' and 'dry' in the appropriate order, the city is mentioned. To say that Tucson is mentioned is to say that it is Tucson's weather, rather than that of Rangoon or Vienna, which is relevant to the truth of the claim. Greetings, such as 'Good Morning' or 'Hello', are examples of speech acts in which no individual is mentioned; their success depends only on the circumstances and the speaker's intentions.

While thinking about the categories of expressions used and individuals mentioned, one should bear in mind that the distinction divides the roles individuals can play rather than the kinds of individuals that can play these roles. That the distinction does not separate individuals by kinds is shown by the fact that linguistic expressions can play both parts: expressions can be mentioned as well as used. By uttering (2), for example, a speaker manages to mention the name 'Tucson' and, in fact, to say something true about it:

(2) The name of the capital city of Arizona contains six letters.

(1) should be compared with (2). The former uses the name of the city to mention the city, while the latter uses a description of the name to mention the name itself.

A detailed understanding of use and mention requires that one draw a distinction between the 'type' of an expression and 'tokens' of that type (see Type/token distinction). Expression tokens are concrete objects, individual marks or sounds made of raw materials: ink, perhaps; or sound waves. Types are abstract individuals and, as such, not made of anything. There is only one word 'Tucson' in the English language (so 'Tucson' is a type), of which there are several tokens in this entry.

The type/token distinction brings with it possibilities for what is used and mentioned. Both tokens and types can be used. To use a token, one has to utter it with the intention of saying something. Using a type requires the same intention, but also requires that the token have the right shape or sound. Tokens and types can also be mentioned. (2) mentions the type and (3) the token used in (1):

(3) The first word token in the first example sentence of the entry on 'Use/mention distinction and quotation' in the *Routledge Encyclopedia of Philosophy* has six letters.

2 Standard uses of quotation and quasi-quotation

Expressions of each grammatical category have what may be called standard uses. The standard use of noun phrases is to mention: 'Tucson', 'The name of the capital city of Arizona' and 'The first word token…*Philosophy*' from (3) above are all used in standard fashion. Expressions of most other categories have standard uses which do not involve mentioning. Articles and quantifiers ('every', 'some', 'all', 'most') do not themselves mention, but rather help determine the range of individuals that are mentioned. Conjunctions govern the manner in which the meanings of compound expressions depend on the meanings of their constituents. According to some theories, adjectives, adverbs and verbs mention relations, properties or events. According to other theories, they do not mention at all.

Along with their standard uses, expressions also have nonstandard uses. Chief among these is their use in quotation. Quotation is a way of using an expression to mention itself. It is the most general and efficient way of talking about expressions. Instead of the long-winded description in (2), a speaker wanting to mention 'Tucson' can simply quote it:

(4) 'Tucson' has six letters.

From the notational point of view, quotation is trivial. Its basic rule is simple: to quote an expression, enclose a token of it in quotation marks - the purpose of the marks being to indicate the nature of the usage. Or it can be simpler yet; one can leave out the marks when it is clear from the context that one wants to talk about an expression rather than use it in the standard way. (Passages which are indented or examples marked by numbers - as in this entry - are quoted using this method.) Quotation is generally used to mention expression types, unless there is an explicit qualification to the contrary.

Quotation is a leveller. No matter what they may standardly be used to mention, when quoted, all nouns mention expressions. It is instructive to compare (1) and (4) on this count to see how quotation can bring about a change in subject matter: (1) is about a large city, while (4), which differs only in the addition of a set of quotation marks, is about a medium-sized word. The effect of quotation on other categories of expressions is even more significant. It reduces the whole variety of standard uses to one. No matter what their standard use may be, when quoted, expressions of all categories become mentioning expressions. This effect can be seen in the example below:

(5) 'A' is an article and 'and' is a conjunction.

In this sentence, the first token of the word 'a' and the second token of the word 'and' are used to mention, while the second token of the word 'a', the first of 'and' and the only one of 'an' operate in standard fashion.

Related to regular quotation is a notation invented by Quine (1940), called 'quasi-quotation'. Quine, who has done more to encourage the correct use of quotation marks than any other writer, created quasi-quotation to help in capturing linguistic generalizations. Quasi-quotation treats an expression in two ways. Towards one part of the expression ('Čamp;' in the example below) it behaves as regular quotation. The other part or parts (' α' and ' β' in the example below) it treats as indeterminate elements that can mention any expression within a specific category.

The marks of quasi-quotation are called corner quotes (' $\ulcorner$ ' on the left and ' $\urcorner$ ' on the right). A full quasi-quotation consisting of the expression enclosed by two corners is itself an indeterminate that can mention any expression consisting of the quoted part and any expression mentioned by the indeterminate part. To take an example from logic, let ' α' and ' β' be indeterminates mentioning sentences and 'Čamp;' be the conjunction symbol. A rule giving the extensional semantics for conjunction can be stated using quasi-quotation:

(6) $\ulcorner \alpha \mathbin{\&} \beta \urcorner$ is true iff α is true and β is true,

where ' $\ulcorner \alpha \mathbin{\&} \beta \urcorner$ ' mentions an expression that is a conjunction of the sentences mentioned by ' α' and ' β'. If α is 'Cicero died' and β is 'Caesar spoke', then $\ulcorner \alpha \mathbin{\&} \beta \urcorner$ is the result of writing α followed by 'Čamp;' followed by β, namely 'Cicero died and Caesar spoke'.

3 Use/mention confusions

The distinction between use and mention can turn slippery. When it does and one loses hold of it, use/mention

confusions can occur. The typical result of such a confusion is that a speaker ends up making a claim about nonlinguistic individuals that is appropriate only to expressions or making a claim about expressions that is only appropriate to nonlinguistic individuals. Where the distinction between use and mention turns most slippery is on the subject of abstract individuals. Numbers and expressions used to mention them can be especially hard to keep apart. The category of numbers includes whole numbers such as 1 and 3, ratios such as $\frac{1}{2}$ and $\frac{4}{3}$ and irrational numbers such as π. The expressions that mention them include the numerals '1' and '3', the fractions ' $\frac{1}{2}$ ' and ' $\frac{4}{3}$ ' and the Greek letter ' π'. Classroom presentations of ratios and fractions commonly involve use/mention confusions. Students are often told something like the following:

Fractions consist of one number divided by another. The number to be divided, called the 'numerator', is placed on the top, while the number to divide by, called the 'denominator', is placed on the bottom, with a line between the two. So, 2 is the numerator and 3 the denominator of $\frac{2}{3}$.

However, the clever student may wonder how this could be. Since $\frac{2}{3}$ is the same quotient as $\frac{4}{6}$, the above remarks suggest that they ought to be the same fraction. Therefore 3 ought to be the denominator of both (on the grounds that the denominators of identicals are identical), but it is not. This puzzle exposes the use/mention confusion. The teacher has confused fractions with ratios. ' $\frac{2}{3}$ ' is a fraction; $\frac{2}{3}$ a ratio. Only fractions have numerators and denominators, while it is the ratios mentioned by fractions that may be identical. The numeral '3' is the denominator of the fraction ' $\frac{2}{3}$ ', but it is the *ratios* $\frac{2}{3}$ and $\frac{4}{6}$ which are identical. Because ' $\frac{2}{3}$ ' is not the same fraction as ' $\frac{4}{6}$ ', they need not, and indeed do not, have the same denominator.

The need to achieve clarity and avoid confusions such as the one above is the main reason why it is important to maintain a clear separation between use and mention. As far as errors go, the mix-up with fractions and ratios is probably mid-level. Some use/mention confusions are of less consequence. Others pose greater problems (see Modal operators §1). Where subtlety and precision are at a premium, as they are in logic and linguistics, use/mention confusions can severely impair the intelligibility of a discussion.

See also: De re/de dicto; Logical and mathematical terms, glossary of

COREY WASHINGTON

References and further reading

Davidson, D. (1984) 'Quotation', in *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press, 1991, 79-92.(Davidson puts forward the view that quotation marks are demonstratives, an alternative to the view advanced in §2.)

Frege, G. (1892) 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25-50; trans. 'On Sense and Reference', in A.P. Martinich (ed.) *The Philosophy of Language*, New York: Oxford University Press, 3rd edn, 1996.(Frege's discussion in this essay contains what may be the first application of the use/mention distinction in the context of a philosophical argument. Frege also introduces the view, stated in §2, that quotation marks are context markers and that a quoted word is used to mention itself.)

Garcia-Carpentera, M. (1994) 'Ostensive Signs: Against the Identity Theory of Quotation', *Journal of Philosophy* 91: 253-64.(Endorses Davidson's theory and defends it against criticisms.)

Geach, P. (1957) *Mental Acts*, London: Routledge & Kegan Paul, 79-82.(Clearest statement of the view, first formulated by Tarski, that quotations can be seen as descriptions.)

Quine, W.V. (1940) *Mathematical Logic*, New York: Norton; repr. Cambridge, MA: Harvard University Press, 1951, 23-37.(A careful discussion of the use/mention distinction, quotation and quasi-quotation.)

Russell, B.A.W. (1919) 'Descriptions', in *Introduction to Mathematical Philosophy*, London: Allen & Unwin, 167-80; repr. in A.P. Martinich (ed.) *The Philosophy of Language*, New York: Oxford University Press, 1985, 213-19; 3rd edn, 1996.(Contains numerous examples of use/mention confusions.)

Tarski, A. (1933) *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw; trans. J.H. Woodger (1956), 'On the Concept of Truth in Formalized Languages', in *Logic, Semantics, Metamathematics*, ed. J. Corcoran, Indianapolis, IN: Hackett Publishing Company, 2nd edn, 1983, 152-278.(The classic work on the view that quotations are names.)

Washington, C. (1992) 'The Identity Theory of Quotation', *Journal of Philosophy* 91: 582-605.(Gives a detailed

description and defence of Frege's view.)