APHASIOLOGY, 2004, *18* (2), 77–102

Spoken word to picture matching from PALPA: A critique and some new matched sets

Jennifer Cole-Virtue and Lyndsey Nickels
Macquarie University, Sydney, Australia

Background: PALPA (Psycholinguistic Assessments of Language Processing in Aphasia; Kay, Lesser, & Coltheart, 1992) is a widely used clinical and research tool. Subtest 47, Spoken word–picture matching, requires the individual with aphasia to listen to a spoken word and correctly choose from five distractor pictures (target, close semantic, distant semantic, visually related, and semantically unrelated). It contributes diagnostically to the clinical evaluation of semantic processing. The authors claim that, first, errors on this test indicate that a semantic comprehension problem is present, and second, that distractor choice reflects the semantic specificity of the problem. For accurate clinical assessment the validity of these claims must be evaluated.

Aims: This paper aims to evaluate the internal validity of PALPA spoken word–picture matching. It addresses two questions; first, is the relationship between the target and distractors what the authors claim it to be? Second, what is the relationship between the target and distractor stimuli in relation to a number of psycholinguistic variables? In addition it allows the clinician to examine the effects of individual variables on performance by including matched subsets of stimuli from this test (matched across five psycholinguistic variables: frequency, imageability, number of phonemes, semantic and visual similarity, word association).

Methods and Procedures: Target and distractor relationships were investigated (in terms of semantic and visual similarity and word category) and psycholinguistic variables (including word frequency, word association, imageability, number of phonemes, semantic and visual similarity).

Outcomes and Results: Analysis revealed a number of confounds within this test: close semantic distractors were not only more semantically similar but also more visually similar to their targets than distant semantic distractors; the semantic and visual (SV) close semantic distractors were more semantically similar to their targets than the non-SV close semantic distractors; targets and distractors did not bear a consistent categorical relationship to their targets, and there were significant intercorrelations between variables for these stimuli (e.g., frequency and length; semantic/visual similarity and length).

Conclusions: The authors' claim that this test assesses semantic comprehension is certainly still tenable. Individuals making errors on this test have a high probability of some semantic processing deficit. However, this study shows that the test fails to assess the *nature* of the semantic processing deficit, as error patterns are subject to the effect of confounding factors. In its current form clinicians should exercise caution when interpreting test findings and be

Address correspondence to: Jennifer Cole-Virtue or Lyndsey Nickels, Macquarie Centre for Cognitive Science (MACCS), Division of Linguistics and Psychology, Macquarie University, Sydney, NSW 2109, Australia. Email: jenny@maccs.mq.edu.au or lyndsey@maccs.mq.edu.au

Thanks to Max Coltheart for helpful comments in the interpretation of these data, and to Stacey Kuan for help in collecting the semantic and visual similarity ratings. Janice Kay made useful comments on an earlier draft. Lyndsey Nickels was funded by an Australian Research Council QEII fellowship during the preparation of this paper.

http://www.tandf.co.uk/journals/pp/02687038.html DOI:10.1080/02687030344000346

aware of its limitations. The development, here, of matched subsets of stimuli allows performance to be re-evaluated in terms of the influence of semantic and visual similarity, imageability, frequency, word length, and word association.

For the practising clinician there are limited cognitive neuropsychological resources for the assessment and treatment of individuals with aphasia. The most widely used assessment materials are published in PALPA (Psycholinguistic Assessments of Language Processing in Aphasia; Kay et al., 1992), which has been an innovative and long-awaited contribution to the clinicians' armoury. PALPA includes a variety of language tasks with assessment and interpretation of findings based on a cognitive neuropsychological approach to language breakdown. Individual tasks within the battery are ''designed to help illuminate the workings of specific components of the language processing model'' (Kay, Lesser, & Coltheart, 1996a, p. 175).

However, there has been relatively little evaluation of the PALPA tasks and the importance of standardisation for the tasks included in the PALPA has been debated in the literature (e.g., Basso, 1996; Ferguson & Armstrong, 1996; Kay et al., 1996a, 1996b). Kay et al. (1992) provide some, but limited, detail regarding normative data, descriptive statistics, and control of variables before each task. Nevertheless, 14 (23%) of the 60 tasks have no normative data or descriptive statistics. Wertz (1996) perceives the lack of standardisation as a problem for the PALPA and comments that this may result in interpretative errors. ''Standardisation provides consistency, validity provides comfort, and reliability provides confidence. The PALPA lacks all three'' (Wertz, 1996, p. 188).

Validity is important, in terms of assessment and treatment planning, because clinicians are reliant on task validity as it ''indicates that the measure does what we think it does'' (Wertz, 1996, p. 184). The authors admit a shortfall and comment, ''we have not carried out psychometrically satisfactory measures of validity or reliability'' (Kay et al., 1996a, p. 160). However, they also respond by suggesting that measures of validity may not be appropriate (Kay et al., 1996b), as validity for PALPA is a question of whether the tasks measure the skills that they claim to and not whether this has validity to external factors, (e.g., anatomical localisation). It is just this question of internal validity that this paper seeks to address for one of the PALPA subtests: Spoken word to picture matching (subtest 47).

Spoken word–picture matching is located in the semantic processing section of PALPA. The aim of this task is to begin to assess semantic comprehension and, in conjunction with other semantic tasks, to enable the clinician to determine if a semantic deficit exists. Clinically, it is one of the most frequently used subtests, perhaps because of the prevalence of semantic processing disorders in aphasia and also that the authors suggest that clinicians use it as a starting point for their assessment of aphasia (Kay et al., 1996a).

As noted above, this task is widely used in speech pathology and neuropsychology as a clinical and research tool to assess language processing skills in people with aphasia. Indeed, in a review of publications citing PALPA, Kay and Terry (2004, this issue) found that it was the most widely used of the PALPA tasks. In both clinical and research settings it is utilised specifically for ''testing semantic ability'' (Marshall, Pound, White-Thomson, & Pring, 1990, p. 174) and to determine if a participant with aphasia has ''a problem in gaining access to semantic information'' (Forde & Humphreys, 1995). It frequently makes a significant diagnostic contribution to the clinicians' quantitative and qualitative evaluation of single word spoken comprehension and is utilised to direct treatment. It is therefore essential to consider whether impaired performance on this task can be interpreted with the confidence that is suggested by its authors, and that the clinical assessment of semantic comprehension ability is accurate.

In order to address this issue, we will investigate the nature of the relationship between target and distractor stimuli in a number of different analyses. We will first focus on the question of whether the relationships are those that Kay et al. (1992) claim in terms of semantic and visual similarity, and superordinate category. Second, we investigate further possible confounds in other psycholinguistic variables that may affect interpretation of performance on this test. Finally, we present some subsets of stimuli from this test that are matched for pertinent psycholinguistic variables.

DESIGN OF THE TARGET AND DISTRACTOR STIMULI

Bishop and Byng (1984) were the first to address the use of different types of semantic distractors in a spoken word–picture match task. Their LUVS Test (The Test for Lexical Understanding with Visual and Semantic Distractors) was designed to assess semantic comprehension ability. They claimed that it was critical to include both semantic and unrelated distractors, to give the subject the opportunity to make semantic errors. Hence, systematic manipulation of distractor type was argued to help define the nature of the comprehension deficit. The PALPA spoken word–picture matching subtest 47 is based on similar principles and was adapted from the original LUVS assessment.

When performing this spoken word–picture matching task, a participant is required to listen to a spoken word and then select the correct picture from a choice of the target and four distractor pictures. There are 40 target items and the distractor pictures for each target consist of ‘‘a close semantic distractor from the same superordinate category, a more distant semantic distractor, a visually similar distractor and an unrelated distractor’’ (Kay et al., 1992, subtest 47, p. 1). For example, for the target word ‘‘carrot’’ the distractor pictures are ‘‘cabbage’’ (close semantic), ‘‘lemon’’ (distant semantic), ‘‘saw’’ (visually related), and ‘‘chisel’’ (unrelated) (see Figure 1). The unrelated and visually related distractors are related to each other semantically but not to the target item. This control feature has been incorporated to prevent the individual responding on the basis of perceived semantic category.

Qualitative evaluation of an individual’s performance involves examination of error type. Kay et al. (1992) state that the distractors have been selected to reflect different semantic relationships with their targets. They claim that the pattern of errors reflects the degree and type of semantic processing impairments: A majority of close semantic errors suggests a relatively high-level semantic impairment. Close semantic distractors are divided into two groups, those that are purely semantically related to their targets (CSDnon-SV, e.g., carrot–cabbage) and those that are semantically *and* visually related to target (CSDSV, e.g., dog–cat). A majority of visually similar semantic errors (CSDSV) is argued to indicate that there may be a perceptual component to the deficit. Perceptual problems are also indicated if the individual tends to choose visually related distractors. The choice of the distant semantic distractor is argued to suggest a more widespread semantic deficit. Lastly, the choice of the unrelated semantic distractor error suggests that there is considerable difficulty in accessing any semantic information regarding the target.

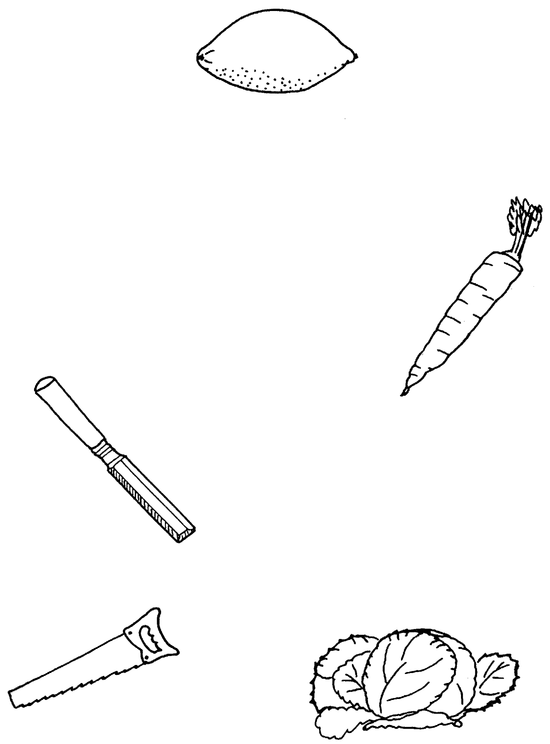

Figure 1. Item 1 from spoken word–picture matching, from PALPA (Kay et al., 1992; subtest 47).

SEMANTIC AND VISUAL SIMILARITY OF TARGET AND DISTRACTORS

The nature of the distractors will be evaluated here in three ways.[1]

(1) Is there a difference in the degree of semantic similarity between targets and the close, distant, and unrelated distractors? Is it the case that there is the predicted gradient of semantic similarity across these items?
(2) Do close and distant semantic distractors differ purely in degree of semantic similarity to their targets, and not differ in visual similarity?
(3) Within the close semantic distractors, do the ''semantic and visual'' stimuli differ from the remaining (purely semantic) close distractors, only in the degree of visual similarity and not in terms of semantic similarity?

Method

To evaluate the relationship between the distractors and the target items, we required a measure of the degree of semantic and visual similarity. Hence, we collected ratings of semantic similarity and visual similarity from 20 Australian non-aphasic participants,

[1] As semantic processing is our focus, here and throughout, we focus on the semantic and unrelated distractors, and do not consider the visual distractor in our discussions.

who were undergraduate psychology students and participated in the experiment as part of the fulfilment of their course requirements or for payment of $10. The participants were asked to judge how semantically similar or visually similar the close semantic distractor, distant semantic distractor, and unrelated distractor were to their corresponding target item. Participants made judgements of either semantic or visual similarity but not both. Participants were asked to use a rating scale of 1–7 to reflect whether words were highly unrelated (0), moderately related, or highly related (7) in meaning or appearance. For semantic similarity, it was emphasised that although some of the word pairs may also be visually similar, the focus for this judgement must be on meaning alone. As semantic processing is our focus, visual distractors were not included in these ratings. All pairs of stimuli were presented as written words in a pseudo-random order so that no target appeared within 10 items of a previous rating of that target.[2]

Results

Ratings of semantic and visual similarity for each item and its close semantic, distant semantic, and unrelated distractors can be found in Tables 1A & 1B.[3]

Semantic similarity across distractor types. Consistent with Kay et al.'s claims, there is a significant difference in semantic similarity across the three distractor types (Page's L test: L = 426.00, $z = -6.97$, $p = .000$). Furthermore, close semantically related distractors are rated as significantly more semantically similar to their targets than both the distant semantic distractors ($t = 8.803$, df 39, $p = .000$), and the unrelated distractors ($t = 24.961$, df 39, $p = .000$). In addition, the distant semantic distractors were rated as significantly more semantically related than the unrelated distractors ($t = 12.012$, df 39, $p = .000$). This semantic similarity gradient (CSD > DSD > URD) is true for 39/40 (97.5%) items. This is not the case for item 35 ''needle'' where the unrelated distractor (tweezers) is rated as more semantically similar to the target than the distant semantic distractor (spinning-wheel). Ideally, all items should consistently show semantic similarity gradients that are in the same direction.

Visual similarity across distractor types. Contrary to the authors' claims that the distractors differ in only semantic similarity, visual similarity was also found to differ significantly across the three distractor types (Page's L test: L = 420.50, $z = -6.67$, $p = .000$). The close semantically related distractors are rated as significantly more visually

[2] For ease of administration we chose to use word pairs, rather than picture pairs. It is possible that this will have resulted in different ratings than if we had used picture pairs. This is likely to have resulted in greater disparities for ratings of visual similarity than of semantic similarity, and for ratings of the visual distractors (not discussed here), than for ratings of other stimuli. However, as it is words that are presented as stimuli to be matched to pictures, we feel it is open to argument which rating is the relevant one for this task (and indeed which may vary from participant to participant). In fact, without knowing how an individual performs the task, and the effects of impairment on this process, it is impossible to adjudicate.

[3] Note that it is likely that the item ''vest'' had a low semantic similarity rating with its target because of dialectal differences between Australian and British English: In Australian English a sleeveless undergarment is called a ''singlet'' while ''vest'' refers to a ''waistcoat''. This, of course, would not be a problem for test administration as ''vest'' is not presented as a word, but simply as a distractor picture. In our opinion, the targets in this test are equally appropriate for Australian participants as for British (although we have received complaints from control participants on this task regarding the item ''hosepipe'' and insisting that 'hose' would be more normal, but this we feel would be equally likely to occur with British participants!).

TABLE 1A
Mean control ratings of semantic and visual similarity for 'non-SV' target and close semantic distractor pairs

Item No.	*Target*	*CSD*	*SV/non-SV*	*Semsim*	*Vissim*	*DSD*	*Semsim*	*Vissim*	*URD*	*Semsim*	*Vissim*
1	carrot	cabbage	non-SV	4.75	2.46	lemon	2.92	1.46	chisel	1.00	3.09
3	hosepipe	bucket	non-SV	4.00	1.60	well	2.33	1.40	frog	1.08	1.20
4	hat	coat	non-SV	4.00	2.18	sock	2.92	1.73	ironing-table	1.17	1.27
6	belt	braces	non-SV	4.00	4.09	shirt	3.08	1.91	clock	1.08	1.18
10	moon	star	non-SV	5.83	4.46	planet	5.17	5.64	anvil	1.13	1.57
12	key	lock	non-SV	5.58	3.55	knob	3.17	2.36	flower	1.18	1.27
13	button	zip	non-SV	5.17	3.09	bow	3.33	2.00	banknote	1.08	1.46
15	syringe	stethoscope	non-SV	3.42	2.00	tablet	2.17	1.36	hinge	1.17	2.27
17	cobweb	spider	non-SV	5.50	2.73	ladybird	1.67	1.64	wagon	1.00	1.36
20	stirrup	saddle	non-SV	5.17	3.40	bridle	5.09	4.63	jacket	2.00	1.33
22	sword	shield	non-SV	4.92	2.00	gun	4.00	3.27	chain	1.92	3.00
23	comb	brush	non-SV	6.83	5.46	mirror	3.58	1.82	ant	1.00	1.46
24	eye	ear	non-SV	5.75	3.18	hair	4.08	1.64	bat	1.33	1.46
27	underpants	vest	non-SV	3.58	3.36	tie	3.17	2.46	watering-can	1.00	1.09
29	paintbrush	palette	non-SV	5.00	2.73	easel	5.46	3.25	kettle	1.42	1.00
32	pram	baby	non-SV	4.67	2.46	teddy	2.92	1.82	towel	1.08	1.55
34	hammock	cot	non-SV	4.75	4.50	pillow	3.25	2.30	cherry	1.33	1.10
35	needle	thimble	non-SV	5.20	3.20	spinning-wheel	2.75	1.91	tweezers	3.17	4.00
37	bell	whistle	non-SV	4.46	3.91	trumpet	3.50	2.18	battery	1.42	1.91
40	stamp	envelope	non-SV	5.83	3.09	pen	2.42	1.82	paint	1.83	1.55

CSD: close semantic distractor; SV: close semantic distractor semantically related to target; non-SV: close semantic distractor semantically related to target; Semsim: mean semantic similarity rating; Vissim: mean visual similarity rating; DSD: distant semantic distractor; URD: unrelated distractor.

TABLE 1B

Mean control ratings of semantic and visual similarity for 'SV' target and close semantic distractor pairs

Item No.	*Target*	*CSD*	*SV/non-SV*	*Semsim*	*Vissim*	*DSD*	*Semsim*	*Vissim*	*URD*	*Semsim*	*Vissim*
2	dog	cat	SV	5.42	5.18	kangaroo	3.25	3.27	butterfly	2.83	1.46
5	axe	hammer	SV	4.75	4.73	scissors	3.58	4.00	kite	1.00	1.36
7	canoe	yacht	SV	5.83	5.09	lifebelt	3.79	1.90	bottle	1.17	1.90
8	ladder	steps	SV	5.33	5.27	rope	3.42	3.00	satchel	2.00	1.46
9	television	radio	SV	5.58	4.55	record-player	4.17	3.64	frying-pan	1.08	1.73
11	apple	orange	SV	6.17	5.73	grapes	4.58	3.18	necklace	1.08	1.09
14	stool	table	SV	4.67	3.09	sofa	5.17	4.18	switch	1.27	1.46
16	crown	tiara	SV	6.56	6.60	gown	1.92	2.18	bread	1.00	1.27
18	candle	match	SV	5.33	4.09	lamp	5.17	4.73	glove	1.17	1.18
19	lobster	crab	SV	6.25	5.46	fish	5.58	4.64	nut	1.83	1.18
21	cow	horse	SV	4.33	5.00	chicken	4.00	2.00	bed	1.00	1.36
25	rake	hoe	SV	5.8	5.33	scarecrow	1.92	2.00	salt	1.00	1.27
26	wall	fence	SV	4.92	5.18	house	4.42	3.09	rocking chair	1.08	1.55
28	nail	screw	SV	5.42	6.55	pliers	3.46	2.36	letter	1.17	1.27
30	parachute	balloon	SV	4.58	5.00	planet	1.00	1.91	puddle	1.00	1.55
31	dart	spear	SV	5.42	5.73	bow	3.50	2.55	razor	1.92	2.82
33	pipe	cigar	SV	5.42	5.00	ashtray	4.08	2.27	rolling-pin	2.42	3.00
36	thumb	finger	SV	6.5	5.64	leg	2.92	2.18	cigarette	1.08	2.64
38	shoe	boot	SV	7.00	6.55	trousers	3.08	2.18	monkey	1.08	1.09
39	mug	cup	SV	6.75	6.91	spoon	3.67	2.64	harp	1.25	1.46
			total mean	5.26	4.25	total mean	3.49	2.61	total mean	1.37	1.66
			SV mean	4.92	3.17	SV mean	3.63	2.90			
			non-SV mean	5.60	5.33	non-SV mean	3.35	2.33			

CSD: close semantic distractor; SV: close semantic distractor semantically related to target; non-SV: close semantic distractor semantically related to target; Semsim: mean semantic similarity rating; Vissim: mean visual similarity rating; DSD: distant semantic distractor; URD: unrelated distractor.

similar to their targets than both the distant semantic distractors (t = 6.408, *df* 39, p = .000) and the unrelated distractors (t = 9.749, *df* 39, p = .000). In addition, the distant semantic distractors were rated as significantly more visually similar to the targets than the unrelated distractors (t = 4.478, *df* 39, p = .000).

If we are to be confident that this test is a measure of semantic processing alone, visual similarity to the target should be held constant across distractors (excluding the visual distractors). As this is clearly not the case, we suggest that the visual similarity differences across distractors represent a confound in this test.

However, within the close semantic distractors, half the items (the CSDSV items) are *designed* to be visually similar to their targets. Hence, for the close semantic distractors, it is only the non-SV items that should be considered in terms of visual similarity. Nevertheless, even within this subset the close semantic distractors are still significantly more visually similar to their targets than the distant (t = 2.566, p = .007) and unrelated (t = 5.238, p = .000) distractors. This finding confirms that the distinction between the close and distant distractors is not one of semantic similarity alone and that another confounding factor, visual similarity, could affect error patterns on these items.

Visual and semantic similarity within close semantic distractors. These analyses compared the close semantic distractors that are claimed to be both semantically and visually related to their targets (CSDSV) with those that are claimed to be purely semantically related (CSDnon-SV). The CSDSV items were indeed rated significantly higher for visual similarity than the CSDnon-SV distractor items (t = 7.300, *df* 38, p = .000). However, the CSDSV items were also rated higher for *semantic* similarity than the CSDnon-SV items (t = 2.668, *df* 38, p = .011). In other words the CSDSV distractors are not only more visually similar but also more semantically similar to their targets than CSDnon-SV distractors.

This finding does not support the authors' premise and has implications for the interpretation of error patterns: A preponderance of errors on CSDsv items does not unequivocally indicate that it is the visual/perceptual component that is responsible for this pattern. As the CSDSV items are also more semantically similar, it would be expected that these items would be more error-prone as the result of a semantic impairment alone. Hence, it is impossible to know whether it is the visual component or the semantic component that results in relatively more CSDSV than CSDnon-SV errors.

SEMANTIC RELATIONSHIPS OF TARGETS AND CLOSE SEMANTIC DISTRACTORS

Kay et al. (1992) make specific claims about the nature of the semantic relationship between the targets and their close semantic distractors. One of these claims is that the target and close semantic distractor pairs share the same superordinate category. For all 40 targets we examine the validity of this claim and seek to define the nature of the semantic relationship between these target pairs.

Method

Twelve participants judged the relationship between the target and the close semantic distractor. They were asked to classify the items as coordinates (that is, from the same semantic category, such as, knife–fork; desk–chair) or semantically associated (things that go together in the world but are not part of the same semantic category; such as, desk–school; flower–vase).

Results and discussion

Of the 40 target and close semantic distractor word pairs, 31 were considered to reflect a coordinate relationship. However, nine target and distractor pairs (hosepipe–bucket, key–lock, syringe–stethoscope, cobweb–spider, candle–match, paintbrush–palette, pram–baby, needle–thimble, stamp–envelope) were consistently classified as semantically associated rather than belonging to the same superordinate category.[4]

Hence, it is not the case that all of the target and distractor pairs share the same semantic relationship. For valid conclusions to be drawn from the analysis of error patterns in this subtest, the semantic relationships between the targets and their distractors need to be constant within each distractor type.

PSYCHOLINGUISTIC PROPERTIES OF STIMULI FROM PALPA SUBTEST 47

In this section we will first examine two variables in detail, word association and word frequency. This will be followed by a brief examination of intercorrelations between a wider set of variables.

Association of targets and distractors

A word association task consists of the subject being asked to say or write the first word that comes to mind for each target word (Lesser, 1981). Word association is thought to be a measure of the lexical relationship between two words, where they may not be semantically linked but often are found to co-occur in the same linguistic context or phrase, e.g., antique vase (Coltheart, 1980). Associative relationships between words have been discussed within the priming literature (see Neely, 1991, for a review). Shelton and Martin (1992) argue that associative priming results from connections between lexical rather than meaning representations. Word association is an important variable to consider in the context of a word–picture matching task, because if it is, in fact, a lexical rather than a featural or semantic measure then it is a potential confounding variable in terms of the semantic claims made for this subtest. Hence, we seek here to establish the nature of the word associations between the targets and distractors in this subtest.

Method

To obtain a measure of the degree of association between the 40 targets and their distractors in subtest 47, the Edinburgh Association Norms were used (EAN; CISD, 1996). The measure of association used is the percentage of participants who produced a particular response to a target. A high percentage response indicates a high degree of association. Of the 40 targets, 5 (12.5%) (hosepipe, lobster, paintbrush, stirrup, underpants) were not found in the EAN and therefore could not be included in the analyses.

[4] It is possible that for some of these pairs a category could be generated to encompass both items. For example, ''medical equipment'' (syringe–stethoscope), ''garden equipment'' (hosepipe–bucket). However, the fact that none of our participants classified these items as belonging to the same category indicates that the categorical relationship is not primary. It is possible that a different design (e.g., tell me the category of these pairs of items) may have resulted in more of these ''less automatic'' categorisations.

Results and discussion

1. Are the targets and semantic distractors associated? The close semantic distractor was produced in response to 24 of the remaining 35 targets (60%). For three of these targets (cobweb, dog, pram) the close semantic distractor was produced by more than 50% of the participants. The distant semantic distractor was produced in response to two (5%) of the targets. For 3 of the 35 targets (ladder, moon, mug) both the close and

TABLE 2
Edinburgh Association Norms for target-close and target-distant semantic distractor pairs

Target	*Close semantic distractor (CSD)*	*% producing CSD*	*Distant semantic distractor (DSD)*	*% producing DSD*
apple	orange	13	grapes	0
axe	hammer	2	scissors	0
bell	whistle	0	trumpet	0
belt	braces	2	shirt	0
button	zip	0	bow	1
candle	match	1	lamp	0
canoe	yacht	1	lifebelt	0
carrot	cabbage	1	lemon	0
cobweb	spider	66	ladybird	0
comb	brush	16	mirror	0
cow	horse	3	chicken	0
crown	tiara	0	gown	0
dart	spear	0	bow	0
dog	cat	57	kangaroo	0
eye	ear	9	hair	0
hammock	cot	0	pillow	0
hat	coat	24	sock	0
hosepipe	bucket	TNF	well	TNF
key	lock	30	knob	0
ladder	steps	12	rope	3
lobster	crab	TNF	fish	TNF
moon	star	2	planet	1
mug	cup	16	spoon	1
nail	screw	1	pliers	0
needle	thimble	0	spinning-wheel	0
paintbrush	palette	TNF	easel	TNF
parachute	balloon	0	plane	3
pipe	cigar	1	ashtray	0
pram	baby	58	teddy	0
rake	hoe	16	scarecrow	0
shoe	boot	4	trousers	0
stamp	envelope	4	pen	0
stirrup	saddle	TNF	bridle	TNF
stool	table	5	sofa	0
sword	shield	7	gun	1
syringe	stethoscope	0	tablet	0
television	radio	14	record-player	0
thumb	finger	2	leg	0
underpants	vest	TNF	tie	TNF
wall	fence	4	house	0

0: Distractor not found; TNF: Target not found distant semantic distractor were found among the subject responses. For the remaining six targets (bell, crown, dart, hammock, needle, and syringe) neither the close or distant semantic distractors were produced in response to the target. See Table 2 for Edinburgh Association Norms for target-close semantic and target-distant semantic distractor pairs.

In summary, 29 of the 35 targets (72.5%) in the database show an association between either their close and/or distant semantic distractor.

2. Does the degree of association vary across the distractors? Ideally, the degree of association between the targets and distractors of different types should be constant, otherwise this presents another confounding factor for this test. However, the percentage of subjects in the EAN who produced the close semantic distractor in response to its target was significantly different from the percentage that produced the distant distractor for the same target (t = 5.535, *df* 34, p = .000). Hence, the degree of association is not constant across the close and distant distractors and may influence distractor choice.

3. What is the relationship between semantic similarity and association? Items that are semantically related can also be highly associated; for example 49% of participants produced the word ''dog'' when given the stimulus ''cat''. If semantic similarity and association were measures of the same relationship then it might be expected that items that are highly associated from the association norms would also be rated highly for semantic similarity. However, for the 35 target items in this subtest where association measures were available, there was no significant correlation between word association ratings and semantic similarity ratings for the same targets. The correlation was, in fact, negligible (r = .051, p = .773). This suggests that measures of semantic similarity and word association are reflecting different aspects of the relationships between items. This is further supported by the fact that Cole-Virtue and Nickels (2004 this issue) found that there was a facilitatory effect on semantic processing for items that had an associative relationship, whereas for semantic similarity, accuracy increased as the semantic similarity rating between items decreased. The opposite directions of these effects are consistent with the idea that measures of association reflect lexical-level relationships and the similarity ratings reflect a more semantically based relationship.

Frequency of targets and distractors

The matching of stimuli for frequency in a semantic task attempts to control for the effect of that variable on that task. According to the Logogen model (Morton, 1970), at a lexical level, the lower the frequency of the item the less likely it is to be accessed correctly, as it has a higher threshold than a high-frequency item. Not only is the frequency of the target important but also the frequency of its neighbours, in this instance the semantic distractors. A low-frequency target is more likely to have neighbours or semantically related distractors that are of a higher frequency. If the semantic system is underspecified, as in aphasia, then this may influence performance on such a task. Hence matching both the target and the distractor for frequency ensures that once activated in the lexicon the reason that a target or its semantic distractor is selected should not be related to its frequency value, but rather reflects the nature of semantic processing.

The role and therefore the need to match items in terms of word frequency in semantic processing tasks is a contentious one. Bishop and Byng (1984) in their test (LUVS), matched distractor stimuli in their spoken word–picture matching task for mean frequency and frequency range. Silveri, Giustolisi, Daniele, and Gainotti (1992) also matched their target and distractor items for word frequency in a spoken word–picture matching task and noted that this was not a variable that affected subject performance. In an investigation of word–picture matching in Alzheimer's disease, Silveri and Leggio (1996) matched target and distractors on frequency range but the performance of the individual with aphasia was not evaluated in terms of this variable.

Kay et al. (1992) do not specify whether targets and distractors in their spoken word–picture matching task are matched for frequency. Here we investigate the relationship between targets and close semantic distractors, first using correlation. We found that the log frequency of targets and the log frequency of the close semantic distractor (CSD) are significantly and positively correlated (r = .51, p = .001). In other words, as the log frequency of the target increases so does the log frequency of the CSD. When we compared the log frequencies of the targets and their corresponding CSD, a t-test confirmed that there was no significant difference between them, $t(39) = 0.20$, 2-tailed, p = 0.84. Similarly, we compared the log frequencies of the targets with their distant semantic distractors (t = 1.142, df 39, p = .26) and log frequencies of the close semantic with the distant semantic distractor (t = 1.015, df 39, p = .31) and found that they were not significantly different. This would suggest that the targets, their close semantic distractors, and distant semantic distractors are matched for log frequency. This is not surprising, as we know that subtest 47 was based on the original LUVS test (Bishop & Byng, 1984) and they matched distractor stimuli for frequency. As a result of the matching, the distractor choice should not be influenced by frequency. However, it is possible that frequency may still affect performance, as the targets themselves consist of higher and lower frequency items.

Intercorrelations between psycholinguistic variables in the stimuli of PALPA subtest 47

Method

Several psycholinguistic variables were examined using correlation. These variables include: spoken word frequency (Celex database; Baayen, Piepenbrock, & Van Rijn, 1993), familiarity, number of syllables, and phonemes (MRC database; Coltheart, 1981). Imageability values for 27 of the target items were taken from the MRC database and a further 8 from a set of object name norms (Morrison, Chappell, & Ellis, 1997). The Morrison et al. set of imageability ratings were linearly transformed so that they could be used in conjunction with those from the MRC database. The same method was used as for the merging of ratings from different sets of data in the MRC database (MRC Psycholinguistic Database User Manual: Version 1; Coltheart, 1981). Semantic and visual similarity ratings for the target–distractor pairs were also used (see earlier). Association norms for 35 of the 40 targets and their distractor pairs were from the Edinburgh Association Norms (CISD, 1996). The number of phonological neighbours for 39/40 of the target items was obtained using a programme written by David Howard (personal communication) to calculate number of phonological neighbours from the Celex database (Baayen et al., 1993).

Results and discussion

A number of variables showed significant correlations within this test (see Appendix A). As expected there are strong correlations between variables that measure similar attributes, such as number of syllables and number of phonemes (r = .84, p = .000), frequency, and familiarity (r = .38, p = .040). It would be expected that words with more syllables would contain more phonemes, and that higher-frequency words are more familiar. In addition, some correlations followed the pattern generally found in the English language. There were significant (negative) correlations between number of phonemes and number of neighbours ($r = -.64$, p = .000) and frequency ($r = -.32$, p = .042): Short words (those with fewer phonemes) have more phonological neighbours (that is, more phonemes in common with other words) and are generally of a higher frequency than long words.

However, other variables, namely number of phonemes, semantic similarity, and visual similarity, show a correlation where a relationship would not necessarily be expected. Both semantic and visual similarity are negatively correlated with number of phonemes (semantic similarity $r = -.358$, p = .023; visual similarity $r = -.321$, p = .043). That is, longer target words are less semantically and visually similar to their distractor words than the shorter target words. This could make the longer target words easier to process, as they are more semantically and visually distinct from their close distractors.

Summary

From the analyses presented above, we have established that there are a number of factors that reduce our confidence in interpreting the results of the PALPA spoken word–picture matching subtest (and will also affect the written word–picture matching subtest, which uses the same stimuli). These factors primarily relate to confounds between variables such as, for example, between semantic and visual similarity, or between word length and semantic similarity. One method that is used to reduce confounds between variables is to develop matched sets of stimuli which control for all relevant variables except the one of interest. Hence if one wished to examine the effects of semantic similarity on performance, two lists of stimuli would be generated which differ in semantic similarity but are identical for all other variables, such as word length, frequency, visual similarity. However, this is no easy exercise, because of the natural intercorrelations between variables in a language (see Cutler, 1981, for a discussion). Below, we present a series of matched sets, which are as well controlled as the limited range of stimuli, and high intercorrelations, allow. Taken together, these sets help overcome some of the limitations of PALPA word–picture matching subtest 47 and extend the generalisations that can be drawn from its use.

Matched sets

As noted above, within subtest 47 we have shown that there are a number of highly intercorrelated variables and the matched sets presented here (Appendices B–F) are designed to control for some of these variables. These will allow for the effects of one variable on performance to be evaluated with confidence. Thus, having tested a person with aphasia on PALPA spoken word–picture matching subtest 47 (or written word–picture matching subtest 48), accuracy on each subset can be calculated. For all of the matched subsets a difference in scores (accuracy or errors) of five or more between the two conditions in a matched subset indicates that the variable being examined does have a significant effect on the individuals' performance. For those clinicians who may not have readily available access to statistical software, significance tables for the matched sets (calculated using Fisher Exact tests) are included in Appendices G–I. Many of these matched sets can, and should, be used in conjunction with one another to determine the nature of the influences on performance. It is important to note, however, that while the presence of a significant difference in accuracy across subsets can be interpreted as evidence for the effects of a variable on performance, the absence of a significant effect does not necessarily imply that there is no such difference—it may be that with a more powerful test with greater numbers of stimuli an effect might emerge.

There are five matched sets, all consisting of subsets of stimuli from PALPA subtest 47:

(1) *Close Semantic Distractor (SV) and Close Semantic Distractor (non-SV) matched for semantic similarity* (Appendix B). This set of 28 items contrasts those close semantic distractors that are classified as semantically *and* visually related to their targets (14 items) with those that are purely semantically related to their target (14 items). While in the test as a whole the close semantic SV and non-SV stimuli differed in semantic similarity as well as in visual similarity, this set is matched for semantic similarity (as well as for frequency, imageability, number of phonemes, and word association). Hence, these subsets differ significantly only in their visual similarity and therefore allow the clinician to determine the effect that visual similarity has on performance.

(2) *High–Low Semantic and Visual Similarity* (Appendix C). This subset contrasts 30 targets that were rated as highly semantically similar (15 items) and lower in semantic similarity (15 items) to their close semantic distractor. They differ significantly in their semantic similarity and are matched for frequency, imageability, and word association. A difference in performance on these conditions allows the clinician to determine the effect of semantic similarity on performance. It has been impossible to match these sets for visual similarity. However, note that the difference in semantic similarity between the high and low conditions is greater than the difference in visual similarity. If the individual shows no effect of visual similarity on the CSDSV/CSDnon-SV set then any difference on this set can be attributed to the effect of semantic similarity.

(3) *High–Low Imageability* (Appendix D). This subset of 30 targets examines the effect of imageability on performance; these are divided into 15 targets rated as high imageability (range 597–637) and 15 as lower in imageability (494–596). Note that the imageability range is necessarily self-restricting, as targets are all picturable items and these tend to be higher in rated imageability. The sets are matched for frequency, semantic similarity, visual similarity, number of phonemes, and word association.

(4) *High–Low Frequency* (Appendix E). This subset includes 28 targets, 14 that are higher frequency and 14 that are lower frequency. A difference in scores between these two conditions allows the clinician to determine the effect of frequency on performance. The sets differ in their frequency values and are matched for semantic similarity, visual similarity, imageability, number of phonemes[5] and word association.

(5) *Number of Phonemes* (Appendix F). This subset consists of 24 target items divided into two sets contrasting target word length. A difference in scores between these two conditions allows the clinician to determine the effect of word length on performance. The sets differ significantly in their word length but are not significantly different in semantic similarity, visual similarity, frequency[6] and imageability.

[5] Note, however, that while there is no significant difference between the number of phonemes in each set, the low-frequency set do tend to be longer. Hence, if an individual shows a strong effect of length on comprehension this could confound performance on this subset.

[6] Note however, that while there is no significant difference between the frequency of each set, the short set do tend to be higher in frequency. Hence, if an individual shows a strong effect of frequency on comprehension this could confound interpretation of length effects on this subset.

GENERAL DISCUSSION

This paper has evaluated spoken word–picture matching from PALPA (subtest 47). There were two main components to this evaluation. First, to evaluate whether the relationship between target and distractors was what Kay et al. (1992) claimed it to be. Second, to evaluate the psycholinguistic properties of the stimuli. We will discuss each of these in turn.

What is the relationship between the target and distractor stimuli in PALPA word–picture matching? (And is the relationship what the authors claim it to be?)

Kay et al. (1992) make three claims about this task: first, that it assesses the semantic processing ability of aphasic participants; second, that the distractors have been selected to reflect differing relationships with the target, consisting of semantically close, distant, visual, and unrelated items; third, that distractor choice or error type will reflect the nature of the semantic processing deficit.

Whether the relationships between the target and distractors fulfil the criteria that the authors' claim has been evaluated using three different measures, semantic similarity, visual similarity, and word category.

Semantic and visual similarity

The basis of the target–distractor relationship is the manipulation of the degree of semantic similarity. However, the authors give no information on how this was achieved or how it was measured, if indeed it was. This study analysed the target–distractor relationships using semantic and visual similarity ratings collected from unimpaired participants. The results and implications of these findings are mixed; the majority challenge the authors' claims regarding the interpretations of error patterns on this test.

In support of this test, the expectation across distractors is that the degree of semantic similarity with the target decreases through close to distant to unrelated distractor items. This was found to be the case, in that the close semantic distractors are more semantically similar than the distant distractors, and the unrelated distractors are less semantically similar to the target than both the close and distant distractors. This supports the authors' premise that there is a gradation of semantic similarity across distractors.

However, other findings contradict the authors' claims. First, although the distractors show a gradation of their semantic similarity relationship, the close distractors are not only more semantically but *also* more visually similar than the distant distractors. This disputes Kay et al.'s claim that the only difference between the close and distant distractor items is the degree of semantic similarity.

Moreover, the close semantic distractors are divided into items that are considered to be semantically and visually related (SV) and those that are purely semantically related (non-SV) to their target. The authors state that errors on the ''SV'' items are suggestive of a visual or perceptual component to an individual's impairment. Hence, the ''SV'' distractors and the ''non-SV'' distractors should differ from each other only on their visual similarity. Analysis showed that, in fact, the ''SV'' and ''non-SV'' distractors differed significantly not only in visual similarity but also in semantic similarity. The implication is that if a subject made a number of errors on ''SV'' distractors, the reason for that error may not always be due just to the visual component of the target–distractor relationship.

These inconsistencies not only put claims regarding the test in dispute, but also mean that the reason for error cannot be distinguished: both a visual impairment and a semantic impairment could influence subject responses.

Word category

Kay et al. (1992, subtest 47, p. 1) state that the target and close semantic distractor items are from the same superordinate category. For the majority of the targets this appears to hold true, but for 22% of targets the relationship was judged to be one of semantic association, either by function or context. For accurate conclusions to be drawn from an individual's performance it is imperative that the relationships between these items is held constant and the authors claims regarding this relationship can only be rejected.

What are the psycholinguistic properties of the stimuli in PALPA word–picture matching?

Word association

Word association ratings are considered to be a measure of the lexical relationship between words. So words can be associated when not necessarily semantically linked. If the semantic similarity and word association ratings were to reflect a similar process, then a significant statistical relationship would be expected between them. This was not found to be the case and means that these variables are, indeed, measuring different aspects of the target–distractor relationship. The majority of the targets were found to be associated with either the close or the distant semantic distractor. This suggests that these items do not differ in the degree of semantic similarity alone and the association between the words may influence distractor choice. If we accept that this is another variable to be considered in the target–distractor relationship, it would be hoped that the degree of association might be held constant across the distractors. Unfortunately, this was not found to be so, as the target/close distractors were significantly more closely associated than the target/distant distractors.

The implication from these findings is that word association is a confound within this test and needs to be considered in the interpretation of performance.

Frequency

There was a positive correlation between the frequency of the targets and the close semantic distractors. This means that as the frequency of the target increases so does the frequency of the close distractor. Further analysis showed that the targets and close and distant distractors are, effectively, matched for frequency providing some control for this variable in distractor choice.

Further variables

Analysis showed that there were many other significant intercorrelations between the variables examined in this study (see earlier). The majority of the correlations were expected and in the predicted directions (e.g., word frequency and word length), however, one unexpected correlation did emerge: longer target words were less semantically and visually similar to their close distractors than shorter target words. The implication is that the longer target words may be advantaged, as they are more semantically and visually distinct from their close distractors.

In sum, there are a number of variables in this test whose possible effects have not been adequately considered in the test design. Many of these may affect performance on this test, therefore making the interpretation of distractor choice and identification of the specific variable that is affecting performance impossible. To help clinicians and researchers overcome these limitations we have presented some subsets of the stimuli that better control for relevant variables such as semantic and visual similarity, imageability, frequency, word association, and word length.[7] Finally, it is important to note that while we have focused on the spoken form of the test (subtest 47) the majority of the findings are directly relevant to the written word–picture matching subtest (subtest 48).

CONCLUSIONS

At the time of publication the PALPA was, and still is in many respects, unique in its contribution to clinical assessment. It has allowed significant progress to be made in the clinical delineation of aphasia, utilising the tools of cognitive neuropsychological enquiry. However, it is now time for its stimuli to be evaluated more closely. This is imperative not only for the clinician, in terms of confidence in and the validity of their assessment, but also for the responsibilities that we have professionally to our clients. This study has attempted to fulfil these aims in two ways—first by evaluating and describing a number of limitations in one subtest of the PALPA, and second, by providing matched sets so that clinicians can address these problems and be confident of the outcomes that this test provides. Lastly, we hope that this is just the beginning: the task of clinical evaluation of the tools that we use is an ongoing process, and the only way is forward.

REFERENCES

Basso, A. (1996). PALPA: An appreciation and a few criticisms. *Aphasiology*, *10*, 190–193.
Baayen, R. H., Piepenbrock, R., & Van Rijn. (1993). *The CELEX lexical database* [CD-ROM]. University of Pennsylvania, PA: Linguistic Data Consortium.
Bishop, D., & Byng, S. (1984). Assessing semantic comprehension: Methodological considerations and a new clinical test. *Cognitive Neuropsychology*, *1*(3), 233–243.
Cole-Virtue, J., & Nickels, L. (2004). Why cabbage and not carrot? An investigation of factors affecting performance on spoken word to picture matching. *Aphasiology*, *18*, 153–179.
Coltheart, M. (1980). The semantic error: Types and theories. In M. Coltheart, K. Patterson, & J. Marshall (Eds.), *Deep dyslexia*. London: Routledge & Kegan Paul.
Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33*A, 497–505.
Coltheart, M., Patterson, K., & Marshall, J. C. (1980). *Deep dyslexia*. London: Routledge & Kegan Paul.
Computing and Information Systems Department (CISD) (1996). *Edinburgh Associative Thesaurus*. Didcot, UK: Rutherford Appleton Laboratory UK.
Cutler, A. (1981) Making up materials is a confounded nuisance, or will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, *10*, 65–70.

[7] An alternative method would be to investigate whether these variables have effects on comprehension for an individual using different stimuli (with greater numbers of items and hence greater sensitivity). If with adequate testing, the variables that confound spoken word–picture matchingt subtest 47 are not found to affect performance, then they are of less concern in the interpretation of performance on this subtest.

Ferguson, F., & Armstrong, E. (1996). The PALPA: A valid investigation of language? *Aphasiology*, *10*(2), 193–197.
Forde, E., & Humphreys, G. W. (1995). Refractory semantics in global aphasia: On semantic organisation and the access–storage distinction in neuropsychology. *Memory*, *3*(3/4), 265–307.
Kay, J., Lesser, R., & Coltheart, M. (1992). PALPA: *Psycholinguistic Assessments of Language Processing in Aphasia*. Hove, UK: Lawrence Erlbaum Associates Ltd.
Kay, J., Lesser, R., & Coltheart, M. (1996a). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology*, *10*(2), 159–179.
Kay, J., Lesser, R., & Coltheart, M. (1996b). PALPA: The proof of the pudding is in the eating. *Aphasiology*, *10*(2), 202–215.
Lesser, R. (1981). *Linguistic investigations of aphasia*. London: Edward Arnold Publications.
Marshall, J., Pound, C., White-Thomson, M., & Pring, T. (1990). The use of picture/word matching tasks to assist word retrieval in aphasic patients. *Aphasiology*, *4*(2), 167–184.
Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50*A (3), 528–559.
Morton, J. (1970). A functional model of memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press.
Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*(6), 1191–1210.
Silveri, M. C., Giustolisi, L., Daniele, A., & Gainotti, G. (1992). Can residual lexical knowledge concern word form rather than word meaning? *Brain and Language*, *43*, 597–612.
Silveri, M. C., & Leggio, M. G. (1996). Influence of disorders of visual perception in word-to-picture matching tasks in patients with Alzheimer's disease. *Brain and Language*, *54*, 326–334.
Wertz, R. T. (1996). The PALPA's proof is in the predicting. *Aphasiology*, *10*(2), 180–190.

APPENDIX A

Intercorrelations between variables for target items in PALPA subtest 47

		LF	*Fam*	*Imag*	*Phones*	*Syll*	*CSD Freq*	*Semsim*	*Vissim*	*AoA*	*Assoc*	*Neigh*
	Mean	0.7	535	594.2	4.18	1.53	0.68	5.26	4.25	264.31	10.60	11.97
	St Dev	0.75	57.17	25.88	1.85	0.72	0.72	0.87	1.43	83.43	17.17	10.20
	Range Min	−0.3	379	494	1	1	−0.3	3.42	1.6	167	0	0
	Range Max	2.45	611	637	9	4	2.42	7	6.91	394	66	0.03
	N	40	29	35	40	40	40	40	40	13	35	39
LF	r		0.38	0.02	−0.32	−0.15	0.52	0.21	0.19	−0.17	0.09	0.45
	sig		**0.040**	ns	**0.042**	ns	**0.001**	ns	ns	ns	ns	**0.004**
Fam	r			0.04	−0.54	−0.30	− 0.21	0.17	0.32	−0.89	−0.47	0.21
	sig			ns	**0.003**	ns	ns	ns	ns	**0.000**	**0.014**	ns
Imag	r				−0.01	0.19	0.12	0.20	0.09	−0.33	0.04	−0.09
	sig				ns	ns	ns	ns	ns	ns	ns	ns
Phon	r					0.84	−0.19	−0.36	−0.32	0.24	0.10	−0.64
	sig					**0.000**	ns	**0.023**	**0.043**	ns	ns	**0.000**
Syll	r						−0.10	−0.24	−0.31	0.01	0.03	−0.60
	sig						ns	ns	ns	ns	ns	**0.000**
CSD Freq	r							0.08	0.06	−0.25	0.20	0.20
	sig							ns	ns	ns	ns	ns
Semsim	r								0.66	−0.40	0.04	0.15
	sig								**0.000**	ns	ns	ns
Vissim	r									−0.16	−0.25	0.23
	sig									ns	ns	ns
AoA	r										0.48	0.28
	sig										ns	ns

LF: log frequency; Fam: familiarity; Imag: imageability; Phones: number of phonemes; Syll: number of syllables; CSD Freq: log frequency of the close semantic distractor; Semsim: semantic similarity rating between target and close semantic distractor; Vissim: visual similarity rating between target and close semantic distractor; AoA: rated word age of acquisition; Assoc: word association; Neigh: number of phonological neighbours.; ns: non significant.

APPENDIX B

Matched set for close semantic distractor-SV and close semantic distractor-non-SV

No. of items	SV set								non-SV set							
	Target	*CSD*	*Semsim*	*Freq*	*Imag*	*Vissim*	*Phones*	*Assoc*	*Target*	*CSD*	*Semsim*	*Freq*	*Imag*	*Vissim*	*Phones*	*Assoc*
1	dog	cat	5.42	1.71	597	5.18	3	57.00	carrot	cabbage	4.75	0.00	602	2.46	5	1.00
2	axe	hammer	4.75	0.00	597	4.73	3	2.00	hat	coat	4.00	1.53	608	2.18	3	24.00
3	canoe	yacht	5.83	−0.30	594	5.09	4	1.00	belt	braces	4.00	1.08	494	4.09	6	2.00
4	ladder	steps	5.33	1.08	612	5.27	4	12.00	moon	star	5.83	1.48	585	4.46	3	2.00
5	television	radio	5.58	2.45	599	4.55	9	14.00	key	lock	5.58	1.57	618	3.55	3	30.00
6	apple	orange	6.17	0.70	637	5.73	4	13.00	button	zip	5.17	1.00	580	3.09	3	0.00
7	stool	table	4.67	0.30	584	3.09	4	5.00	sword	shield	4.92	0.70	597	2.00	3	7.00
8	candle	match	5.33	0.70	577	4.09	6	1.00	comb	brush	6.83	0.60	592	5.46	3	16.00
9	rake	hoe	5.80	0.00	550	5.33	3	16.00	eye	ear	5.75	2.00	603	3.18	1	9.00
10	wall	fence	4.92	1.83	576	5.18	3	4.00	paintbrush	palette	5.00	−0.30	601	2.73	8	
11	nail	screw	5.42	0.95	588	6.55	3	1.00	pram	baby	4.67	0.00	579	2.46	4	58.00
12	parachute	balloon	4.58	0.00	598	5.00	7	0.00	hammock	cot	4.75	0.00	562	4.50	3	0.00
13	dart	spear	5.42	0.00	636	5.73	3	0.00	needle	thimble	5.20	0.60	589	3.20	5	0.00
14	pipe	cigar	5.42	1.43	598	5.00	3	1.00	bell	whistle	4.46	1.69	610	3.91	5	0.00
	mean		5.33	0.78	595.93	5.04	4.21	9.07	mean		5.06	0.85	587.14	3.37	3.93	11.46
	st dev		0.46	0.84	22.61	0.81	1.85	14.96	st dev		0.76	0.74	30.54	1.01	1.73	17.14
									p-value		0.27	0.79	0.39	0.00	0.68	0.70

CSD: close semantic distractor; Semsim: semantic similarity rating between target and close semantic distractor; Freq: log frequency; Imag: rated imageability; Vissim: visual similarity rating between target and close semantic distractor; Phones: number of phonemes; Assoc: word association; *p*-value: significance of a *t*-test comparing SV and non-SV sets.

APPENDIX C

Matched set for high–low semantic and visual similarity

No of items	*High Sem/Vis sim set*						*Low Sem/Vis sim set*					
	Target	*Semsim*	*Freq*	*Imag*	*Vissim*	*Assoc*	*Target*	*Semsim*	*Freq*	*Imag*	*Vissim*	*Assoc*
1	canoe	5.83	−0.30	594	5.09	1.00	carrot	4.75	0.00	602	2.46	1.00
2	television	5.58	2.45	599	4.55	14.00	hosepipe	4.00	−0.30	.	1.60	
3	moon	5.83	1.48	585	4.46	2.00	hat	4.00	1.53	608	2.18	24.00
4	apple	6.17	0.70	637	5.73	13.00	axe	4.75	0.00	597	4.73	2.00
5	key	5.58	1.57	618	3.55	30.00	belt	4.00	1.08	494	4.09	2.00
6	crown	6.56	1.34	602	6.60	0.00	stool	4.67	0.30	584	3.09	5.00
7	cobweb	5.50	0.60	.	2.73	66.00	syringe	3.42	−0.30	596	2.00	0.00
8	lobster	6.25	−0.30	630	5.46		cow	4.33	1.08	632	5.00	3.00
9	comb	6.83	0.60	592	5.46	16.00	sword	4.92	0.70	597	2.00	7.00
10	eye	5.75	2.00	603	3.18	9.00	wall	4.92	1.83	576	5.18	4.00
11	rake	5.80	0.00	550	5.33	16.00	underpants	3.58	0.00	.	3.36	
12	thumb	6.50	0.78	599	5.64	2.00	parachute	4.58	0.00	598	5.00	0.00
13	shoe	7.00	0.78	601	6.55	4.00	pram	4.67	0.00	579	2.46	58.00
14	mug	6.75	0.85	574	6.91	16.00	hammock	4.75	0.00	562	4.50	0.00
15	stamp	5.83	0.90	.	3.09	4.00	bell	4.46	1.69	610	3.91	0.00
	mean	6.12	0.90	598.77	4.95	13.79	mean	4.39	0.51	587.31	3.44	8.15
	st dev	0.50	0.78	22.53	1.33	17.21	st dev	0.48	0.75	33.03	1.27	16.29
							p-value	0.000	0.173	0.312	0.003	0.392

Semsim: semantic similarity rating between target and close semantic distractor; Freq: log frequency; Imag: rated imageability; Vissim: visual similarity rating between target and close semantic distractor; Assoc: word association; *p*-value: significance of a paired-samples *t*-test across sets.

APPENDIX D

Matched set for high-low imageability

No. of items	*High Imag set*							*Low Imag set*						
	Target	*Imag*	*Freq*	*Semsim*	*Vissim*	*Phones*	*Assoc*	*Target*	*Imag*	*Freq*	*Semsim*	*Vissim*	*Phones*	*Assoc*
1	carrot	602	0.00	4.75	2.46	5	1	belt	494	1.08	4.00	4.09	4	2
2	hat	608	1.53	4.00	2.18	3	24	canoe	594	−0.30	5.83	5.09	4	1
3	ladder	612	1.08	5.33	5.27	4	12	moon	585	1.48	5.83	4.46	3	2
4	television	599	2.45	5.58	4.55	9	14	button	580	1.00	5.17	3.09	5	0
5	apple	637	0.70	6.17	5.73	4	13	stool	584	0.30	4.67	3.09	4	5
6	key	618	1.57	5.58	3.55	2	30	syringe	596	−0.30	3.42	2.00	6	0
7	crown	602	1.34	6.56	6.60	4	0	candle	577	0.70	5.33	4.09	6	1
8	lobster	630	−0.30	6.25	5.46	6		comb	592	0.60	6.83	5.46	3	16
9	cow	632	1.08	4.33	5.00	2	3	rake	550	0.00	5.80	5.33	3	16
10	eye	603	2.00	5.75	3.18	1	9	wall	576	1.83	4.92	5.18	3	4
11	paintbrush	601	−0.30	5.00	2.73	8		nail	588	0.95	5.42	6.55	3	1
12	dart	636	0.00	5.42	5.73	3	0	pram	579	0.00	4.67	2.46	4	58
13	thumb	599	0.78	6.50	5.64	3	2	hammock	562	0.00	4.75	4.50	5	0
14	bell	610	1.69	4.46	3.91	3	0	needle	589	0.60	5.20	3.20	5	0
15	shoe	601	0.78	7.00	6.55	2	4	mug	574	0.85	6.75	6.91	3	16
	mean	612.67	0.96	5.51	4.57	3.93	8.62	mean	574.67	0.59	5.24	4.37	4.07	8.13
	st dev	14.23	0.84	0.89	1.47	2.25	9.71	st dev	25.41	0.64	0.92	1.43	1.10	15.10
								p-value	0.000	0.180	0.414	0.707	0.838	0.922

Imag: rated imageability; Freq: log frequency; Semsim: semantic similarity rating between target and close semantic distractor; Vissim: visual similarity rating between target and close semantic distractor; Phones: number of phonemes; Assoc: word association; *p*-value: significance of a paired-samples *t*-test across sets.

APPENDIX E

Matched set for frequency

No. of items	High Freq set									Low Freq set								
	Target	*CDSV*	*Freq*	*Celex*	*Semsim*	*Vissim*	*Imag*	*Phones*	*Assoc*	*Target*	*CD non-SV*	*Freq*	*Celex*	*Semsim*	*Vissim*	*Imag*	*Phones*	*Assoc*
1	dog	cat	1.71	51.00	5.42	5.18	597	3	57	carrot	cabbage	0.00	1.00	4.75	2.46	602	5	1
2	hat	coat	1.53	34.00	4.00	2.18	608	3	24	axe	hammer	0.00	1.00	4.75	4.73	597	3	2
3	belt	braces	1.08	12.00	4.00	4.09	494	4	2	canoe	yacht	−0.30	0.50	5.83	5.09	594	4	1
4	ladder	steps	1.08	12.00	5.33	5.27	612	4	12	stool	table	0.30	2.00	4.67	3.09	584	4	5
5	television	radio	2.45	285.00	5.58	4.55	599	9	14	syringe	stethoscope	−0.30	0.50	3.42	2.00	596	6	0
6	moon	star	1.48	30.00	5.83	4.46	585	3	2	lobster	crab	−0.30	0.50	6.25	5.46	630	6	
7	key	lock	1.57	37.00	5.58	3.55	618	2	30	comb	brush	0.60	4.00	6.83	5.46	592	3	16
8	button	zip	1.00	10.00	5.17	3.09	580	5	0	rake	hoe	0.00	1.00	5.80	5.33	550	3	16
9	crown	tiara	1.34	22.00	6.56	6.60	602	4	0	paintbrush	palette	−0.30	0.50	5.00	2.73	601	8	
10	cow	horse	1.08	12.00	4.33	5.00	632	2	3	parachute	balloon	0.00	1.00	4.58	5.00	598	7	0
11	wall	fence	1.83	68.00	4.92	5.18	576	3	4	dart	spear	0.00	1.00	5.42	5.73	636	3	0
12	nail	screw	0.95	9.00	5.42	6.55	588	3	1	pram	baby	0.00	1.00	4.67	2.46	579	4	58
13	pipe	cigar	1.43	27.00	5.42	5.00	598	3	1	hammock	cot	0.00	1.00	4.75	4.50	562	5	0
14	bell	whistle	1.69	49.00	4.46	3.91	610	3	0	needle	thimble	0.60	4.00	5.20	3.20	589	5	0
		mean	1.44	47.00	5.14	4.61	592.79	3.64	10.71		mean	0.02	1.36	5.14	4.09	593.57	4.71	8.25
		st dev	0.41	70.81	0.73	1.22	32.26	1.74	16.39		st dev	0.30	1.18	0.85	1.35	22.38	1.59	16.76
											p-value	0.000	0.023	0.982	0.288	0.941	0.101	0.709

Freq: log frequency; Celex: spoken word frequency values from Celex database; Semsim: semantic similarity rating between target and close semantic distractor; Vissim: visual similarity rating between target and close semantic distractor; Imag: rated imageability; Phones: number of phonemes; Assoc: word association; *p*-value: significance of a paired-samples *t*-test across sets.

APPENDIX F

Matched set for number of phonemes

No. of items	Long set						Short set					
	Target	*Phones*	*Semsim*	*Freq*	*Imag*	*Vissim*	*Target*	*Phones*	*Semsim*	*Freq*	*Imag*	*Vissim*
1	needle	5	5.20	0.60	589	3.20	cow	2	4.33	1.08	632	5.00
2	button	5	5.17	1.00	580	3.09	key	2	5.58	1.57	618	3.55
3	candle	6	5.33	0.70	577	4.09	axe	3	4.75	0.00	597	4.73
4	syringe	6	3.42	−0.30	596	2.00	bell	3	4.46	1.69	610	3.91
5	lobster	6	6.25	−0.30	630	5.46	comb	3	6.83	0.60	592	5.46
6	canoe	4	5.83	−0.30	594	5.09	dart	3	5.42	0.00	636	5.73
7	parachute	7	4.58	0.00	598	5.00	dog	3	5.42	1.71	597	5.18
8	paintbrush	8	5.00	−0.30	601	2.73	hat	3	4.00	1.53	608	2.18
9	television	9	5.58	2.45	599	4.55	moon	3	5.83	1.48	585	4.46
10	belt	4	4.00	1.08	494	4.09	eye	1	5.75	2.00	603	3.18
11	carrot	5	4.75	0.00	602	2.46	pram	4	4.67	0.00	579	2.46
12	hammock	5	4.75	0.00	562	4.50	rake	3	5.80	0.00	550	5.33
	mean	5.83	4.99	0.39	585.17	3.85	mean	2.75	5.24	0.97	600.58	4.26
	st dev	1.53	0.77	0.83	33.10	1.13	st dev	0.75	0.81	0.80	23.47	1.20
							p-value	0.000	0.451	0.092	0.202	0.399

Phones: number of phonemes; Semsim: semantic similarity rating between target and close semantic distractor; Freq: log frequency; Imag: rated imageability; Vissim: visual similarity rating between target and close semantic distractor; *p*-value: significance of a paired-samples *t*-test across sets.

APPENDIX G

Significance table for hi/low semantic/visual similarity set and imageability matched sets where $n = 15$

Scores	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
0	ns	.042	.017	.006	.002	.001	.000	.000	.000	.000	.000	.000
1		ns	ns	.035	.014	.005	.002	.001	.000	.000	.000	.000
2				ns	.050	.021	.008	.003	.001	.000	.000	.000
3					ns	ns	.025	.009	.003	.001	.000	.000
4							ns	.027	.009	.003	.001	.000
5								ns	.025	.008	.002	.000
6									ns	.021	.005	.001
7										.050	.014	.002
8										ns	.035	.006
9											ns	.017
10												.042
11												ns

Any combination of scores that does not appear on the table is nonsignificant (e.g., 0 & 2; 12 & 14).

APPENDIX H

Significance table for CSDSV/CSDnon-SV and frequency matched sets where $n = 14$

Scores	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*
0	ns	.041	.016	.006	.002	.001	.000	.000	.000	.000	.000
1		ns	ns	.033	.013	.004	.001	.000	.000	.000	.000
2		ns	ns	ns	.046	.018	.006	.002	.000	.000	.000
3				ns	ns	ns	.021	.007	.002	.000	.000
4						ns	ns	.021	.006	.001	.000
5								ns	.018	.004	.001
6									.046	.013	.002
7									ns	.033	.006
8										ns	.016
9											.041
10											ns

APPENDIX I

Significance table for number of phonemes matched set where $n = 12$

Scores	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
0	ns	.035	.012	.004	.001	.000	.000	.000	.000
1		ns	ns	.024	.008	.002	.000	.000	.000
2				ns	.030	.009	.002	.000	.000
3					ns	.030	.008	.001	.000
4						ns	.024	.004	.001
5							ns	.012	.005
6								.035	.014
7								ns	.037
8									ns