Syntax
A Journal of Theoretical, Experimental and Interdisciplinary Research

Syntax 15:4, December 2012, 354–396 DOI: 10.1111/j.1467-9612.2012.00172.x

Catenae: Introducing a Novel Unit of Syntactic Analysis

Timothy Osborne, Michael Putnam, and Thomas Groß

Abstract. This paper introduces a novel unit of syntactic analysis, the *catena* (Latin for 'chain'; plural *catenae*). The catena is defined in a dependency-based grammar as a word or a combination of words that is continuous with respect to dominance. According to this definition, any dependency tree or any subtree (complete or partial) of a dependency tree qualifies as a catena. The paper demonstrates that idioms are stored as catenae and that the elided material of ellipsis mechanisms (e.g., answer fragments, gapping, stripping, VP ellipsis, pseudogapping, sluicing, and comparative deletion) is a catena. Constituents are always catenae, but many catenae are *not* constituents. Based on the flexibility and utility of the catena concept, the claim is put forth and defended that the catena is the fundamental unit of syntax, not the constituent.

1. Introduction

This paper is about a novel unit of syntactic analysis, the *catena* (Latin for 'chain'). The catena (plural catenae) is defined in a dependency-based grammar as a word or a combination of words that is continuous with respect to dominance. O'Grady (1998) introduced the catena concept—although he employed the term *chain* instead of *catena*—as the basis for his account of the syntax of idioms.[1] This paper extends the catena concept to ellipsis, demonstrating that the elided material of many ellipsis mechanisms (e.g., answer fragments, gapping, pseudogapping, VP ellipsis, pseudogapping, sluicing, and comparative deletion) is always a catena but not always a constituent. The stance put forth and defended in this paper is that the catena is the basic unit of syntax, not the constituent as understood in constituency grammars (and dependency grammars).

As a point of departure, consider the following sentences. They illustrate the difficulties facing theories of syntax that assume the constituent to be the fundamental unit:

(1) She *has lost* her keys.

(2) Fred *took* us *on*.

(3) We *wiped the floor with* them.

[1] We too use the term *chain* in many of our earlier works (Osborne 2005, 2006, 2007, 2008; Groß & Osborne 2009; Groß 2010). The term *catena* (in place of *chain*) is introduced in this paper to avoid confusion. A chain in many theories is understood as a derivational legacy consisting of a head and one or more traces or copies of that head lower in the structure. The catena concept has little to do with these chains. Furthermore, in set theory, a chain is a totally ordered set. A catena, in contrast, can be a set of words that is either totally or only partially ordered with respect to dominance.

(4) *What is* that fly *doing* in my soup?

(5) Larry will *persuade* you *to stay* sooner than
Susan will me.

(6) She has *more old pictures* of you than
he has of me.

The discussion focuses on the words in italics. The word combination *has lost* is the matrix predicate in (1), and as such, it is a semantic unit, yet this semantic unit is not a constituent in surface syntax (at least not in most modern theories). Similarly, the phrasal verb *took . . . on* is clearly a semantic unit in (2), yet not a constituent in surface syntax, because the pronoun *us* intervenes. The words *wiped the floor with* constitute the idiom in (3) and are thus a semantic unit, yet this unit cannot be seen as a constituent because it does not include the object of the preposition. The word combination *What is . . . doing* takes on the special meaning 'why' in (4), but this unit does not form a constituent to the exclusion of the subject. The pseudogapped words *persuade . . . to stay* in (5) and the elided words *more old pictures* of comparative deletion in (6) do not qualify as constituents, yet they must qualify as some sort of unit of syntax, given that the gaps of pseudogapping and the elided words of comparative deletion are not arbitrary.

In general, these cases and others like them contain either noncontiguous sequences of words that cannot be analyzed as surface constituents, as in (2), (4), and (5), or contiguous sequences that are not typically analyzed as constituents, as in (1), (3), and (6). In these regards, this paper demonstrates that each of the word combinations in italics in (1–6) qualifies as a catena in dependency grammar.

Three claims form the core argument of this paper.

Claim 1: The catena is the fundamental unit of syntax, not the constituent.

Claim 2: All idioms are stored as catenae but not all idioms are stored as constituents.

Claim 3: The elided material of many ellipsis mechanisms (answer fragments, gapping, stripping, VP ellipsis, pseudogapping, sluicing, comparative deletion) is always a catena, but not always a constituent.

Claims 2 and 3 are in a sense derived from claim 1. The discussion focuses on data from English, although several examples from German are also used to solidify certain points. In this regard however, we believe that the catena concept is applicable to other languages and that its applicability across languages is a major strength of the concept.

To conclude this introduction, an important point concerning the use of dependency grammar (as opposed to constituency grammar) should be mentioned. The catena concept has been discovered and is being developed within a dependency-based grammar. An anonymous reviewer emphasizes, however, that the catena can also be defined over constituency-based structures. Indeed, for constituency grammars the catena can be defined as *a word or a combination of words the projections of which are continuous with respect to dominance*. The discussion in this

paper does not, however, consider the catena from a constituency-grammar point of view; rather, it focuses on its justification and utility in our dependency-based system. The fact that the concept can be defined over both grammar types is, however, another general strength of the concept and important for the validity of claim 1.[2]

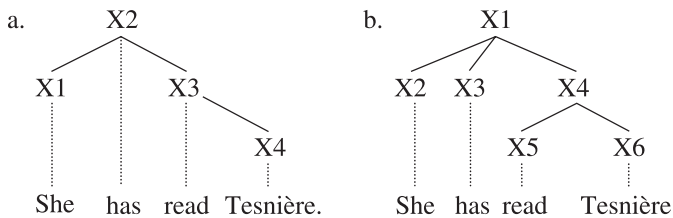
2. Dependency Grammar

The following four subsections present some central aspects of the current dependency grammar. These aspects are consistent in many respects with an established tradition (e.g., Tesnière 1959; Hays 1964; Robinson 1970; Kunze 1975; Matthews 1981; Sgall, Hajičová & Panevová 1986; Mel'čuk 1988; Schubert 1988; Starosta 1988; Lobin 1993; Pickering & Barry 1993; Engel 1994; Jung 1995; Heringer 1996; Groß 1999; Eroms 1985, 2000; Kahane 2000; Tarvainen 2000; Ágel et al. 2003–2006).

2.1 One-to-One

A very noticeable difference between dependency-based and constituency-based theories of syntax lies in the word-to-node ratio. Dependency is a one-to-one relation, whereas constituency is a one-to-one-or-more relation. For every word in a sentence, there is exactly one node in the dependency structure thereof. The corresponding constituency structure, in contrast, has many words in the sentence corresponding to more than one node in the structure.

(7)



The constituency tree (7b) is the direct translation (dependency → constituency) of the dependency tree (7a). Note in this regard that (7b) lacks a finite VP constituent.[3]

[2] Although the catena can indeed be defined over constituency-based structures, we believe that the concept is insightful only for those constituency-based grammars that are representational, allow strictly endocentric structures only, and acknowledge *n*-ary branching. For those derivational constituency grammars that allow binary branching only, the applicability of the catenae concept is debatable, because at times empty functional categories (occupied by traces or copies) will intervene in the hierarchy and obscure the word combinations that qualify as catenae.

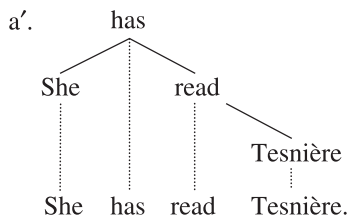
[3] Most (if not all) dependency grammars do not acknowledge a finite VP constituent. They do, however, readily acknowledge nonfinite VP as a constituent. In fact, this is a misunderstanding that dependency grammars grapple with. One inaccurately assumes that because dependency grammars do not acknowledge a finite VP constituent, they therefore reject VP as a constituent in general. The reader is invited to check the dependency trees in this article in this regard. Finite VP is not ever shown as a constituent, whereas numerous trees presented here contain nonfinite VP constituents.

Dependency is incapable of acknowledging the initial binary division (e.g., S → NP + VP) that is central for most constituency grammars.[4]

There are four words in the sentence and exactly four nodes in the dependency structure in (7a). The strict mother–daughter relation of dependency prohibits the nodes from outnumbering the words. The constituency structure in (7b), in contrast, has six nodes in the structure. The part–whole relation of constituency necessitates that the number of nodes in the structure always outnumber the number of words (by at least one). Trees (7a,b) thus demonstrate that dependency trees are truly minimal in comparison with their constituency tree counterparts.[5]

The one-to-one relation inherent to dependency allows one to dispense with the word-node distinction entirely. One puts the words themselves directly into the hierarchy in the following manner:

(7)



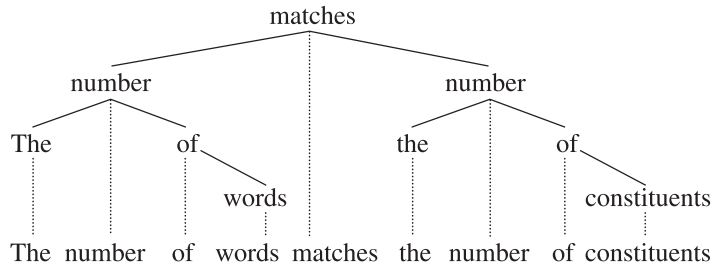
Given that this convention results in particularly transparent trees, it shall be used throughout this article. A word is a node, and a node is a word.

Dependency is like constituency insofar both relations group words into constituents. The constituent is defined in a neutral manner as *a word/node plus all the words/nodes that that word/node dominates*. The number of constituents in a given structure is equivalent to the number of words/nodes:

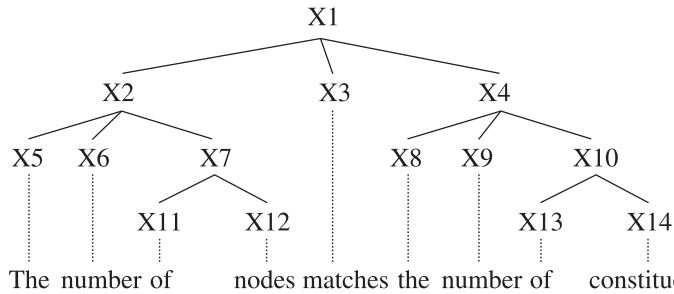
[4] On a number of occasions in this article, the dependency tree and the corresponding constituency tree are placed adjacent to each other (or the one immediately below the other). This practice facilitates comparison. One must be aware, however, that the constituency trees produced, given that they are direct translations of their dependency-tree counterparts, are unlike the constituency trees that one typically finds in the established constituency-based frameworks. They are flatter than usual.

[5] It is an oversimplification (and at times wrong) to assert that simpler structures are better. But if simpler structures succeed to the same extent as more complex structures at accounting for given phenomena, then the simpler structures *are* better. This is Occam's Razor. The most vivid example of the efficacy of the simpler dependency structures is with constituency tests in English. Osborne (2005:254ff., 2006:53ff., 2008:1126ff.) demonstrates that standard constituency tests (e.g., topicalization, clefting, pseudoclefting, pro-form substitution, and answer fragments) identify much less structure than most constituency grammars posit. The minimal dependency structures are much more accurate in this area.

(8) a.



b.



The constituency tree here is again a direct translation of the dependency tree. A finite VP is therefore absent, and the NPs are flatter than many constituency grammars assume. Given that there are nine words in the dependency tree in (8a), there are nine constituents there. In the constituency tree in (8b) in contrast, there are 14 nodes, so there are 14 constituents.

Both views of constituent structure assume that the words are *not* ordered in an arbitrary manner; rather, they are grouped. For instance, both trees show the subject NP *the number of words* and the object NP *the number of constituents* as complete subtrees (= constituents). In this regard, each tree should have a structure that matches best the results of standard constituency tests (e.g., topicalization, clefting, pseudoclefting, pro-form substitution, and answer fragments). These tests confirm, for instance, that the subject NP and the object NP are indeed constituents as shown in the trees.

2.2 Units of Syntax

The current dependency grammar uses distinct terminology to denote units of syntax. This terminology is possible based on three primitive assumptions: (1) units of syntax are organized with respect to precedence; (2) units of syntax are also organized with respect to dominance; and (3) dominance is manifest as the one-to-one dependency relation.

Given these primitive assumptions, the fundamental units of the syntax are defined as follows:

String: A word or a combination of words that is continuous with respect to precedence

Catena: A word or a combination of words that is continuous with respect to dominance

Constituent: A catena that consists of a word plus all the words that that word dominates

Root: The one word in a given catena that is not dominated by any other word in that catena

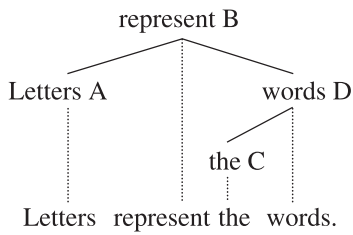
Head: The one word that immediately dominates a given catena

Dependent: A constituent that is immediately dominated by a given word

Governor: The one word that licenses the appearance of a given catena

These units are illustrated using the following tree:

(9)



The capital letters serve to abbreviate the words.

Strings and catenae: The string is defined purely in terms of the *x*-axis, that is, with respect to precedence alone. There are 10 distinct strings in (9): A, B, C, D, AB, BC, CD, ABC, BCD, and ABCD. The catena is defined purely in terms of the *y*-axis, that is, with respect to dominance alone. In graph-theoretic terms, a catena is any tree or any subtree of a tree.[6] There are 10 distinct catenae in (9): A, B, C, D, AB, BD, CD, ABD, BCD, and ABCD.[7] Notice that a catena can have two or more branches, such as ABD and ABCD. Notice also that some strings are not catenae, such as BC and ABC, and some catenae are not strings, such as BD and ABD. There are five noncatena word combinations: AC, AD, BC, ABC, and ADC. Section 3 has more to say about the word combinations that qualify as catenae.

Constituents and dependents: A constituent is a particular type of catena—namely, one that is complete (= complete subtree). There are four distinct constituents in (9): A, C, CD, and ABCD. The number of constituents in a given tree is always equivalent to the number of words in that tree. This fact holds because of the one-to-one dependency relation, whereby a word is a node and a node is a word.

[6] This fact suggests that the term *subtree* instead of *catena* would be a better term to denote the intended word combinations. The problem with the term *subtree* is that it would not be applicable to constituency structures. The word combinations that qualify as subtrees of dependency trees can hardly be understood as subtrees in the corresponding constituency trees.

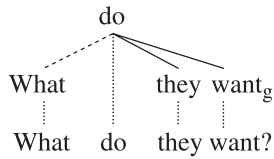
[7] When listing and discussing catenae in this section and section 3, the elements (= words) that form catenae are always listed in their left-to-right order. This convention simplifies the counting of catenae as the trees become larger, as will become evident in section 3.

A dependent is a constituent that has a given node/word as its head. For instance, the dependents of B in (9) are A and CD, and the dependent of D is C.

Roots and heads: The root of a given catena is the one word in that catena that is not dominated by any other word in that catena, whereas the head of a given catena is the one word outside of that catena that immediately dominates that catena. Take the catena CD as an example: its root is D and its head is B. Take the catena ABD as a second example: its root is B and it has no head.

Governors: The governor of a given catena is the one word that licenses the appearance of that catena. The governor and the head of a given catena are usually one and the same word. When a discontinuity is perceived, however, the governor of the relevant catena is not its head.

(10)



The head of *what* is *do* because *do* immediately dominates *what*. The governor of *what*, however, is *want* because *want* subcategorizes for an object. Those catenae whose heads are not their governors are indicated via the dashed dependency edge. In such cases, *rising* is assumed. The governor of the risen catena is indicated with the *g* subscript. Rising is examined in the next sections.

To conclude this section, some comments about the distinction between catenae and constituents are necessary. The constituent is the basic unit of constituency grammars. In this respect, most dependency grammars do not employ the term *constituent*. These grammars do, however, have terms to denote a complete subtree of a tree (= constituent), although this terminology varies (see, e.g., Tesnière 1969:14, Hays 1964:520, Mel'čuk 1988:14, Groß 1999:69, Eroms 2000:86ff., Hudson 1984:92, Siewierska 1988:142, Hellwig 2003:603). The current dependency grammar uses the term *constituent* to denote a complete subtree (= a word/node plus all the words/nodes that that word/node dominates). This use of terminology is advantageous because it makes a comparison of constituent structure possible across dependency and constituency grammars. The thing that one must understand about catenae and constituents is that a constituent (be it a dependency-grammar constituent or a constituency-grammar constituent) is a subtype of catena. Thus a constituent is always a catena, but there are many catenae that are not constituents.

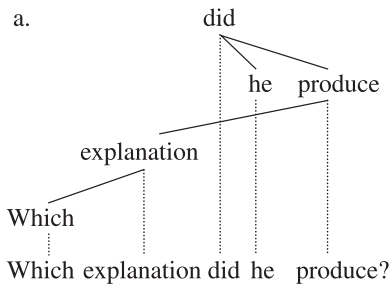
2.3 Rising

Dependency structures are typically less layered than constituency structures. The flatter structures give rise to fewer discontinuities. Despite this fact, discontinuities are a common occurrence, and it is therefore necessary that the theory have a means of addressing them. The current system addresses discontinuities in terms of rising, following Osborne (2007:34ff.) and Groß & Osborne (2009). Rising

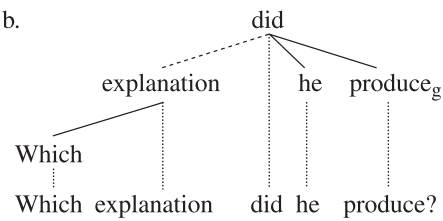
denotes a constellation where the governor of a given catena is not the head of that catena.

The following dependency trees illustrate three types of rising. A German example is used to illustrate scrambling rising.[8]

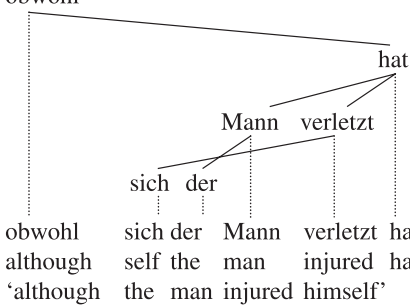
(11) a.



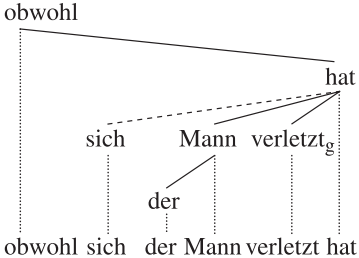
b.



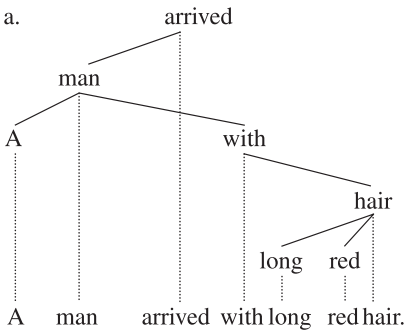
(12) a. obwohl



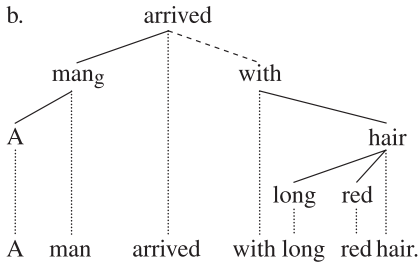
b. obwohl



(13) a.



b.



The trees on the left contain discontinuities, as indicated by the crossing lines. The corresponding trees on the right show how such perceived discontinuities are

[8] Examples (11)–(13) show the determiners and quantifiers as dependents of their nouns. They are consistent in this regard with most dependency grammars, which assume NP (not DP) for noun phrases. The issue is important for the overall theory of catenae. Numerous idioms include an object noun but exclude that noun's possessive (e.g., *pull X's leg*, *save X's bacon*, *tread on X's toes*). This fact supports NP over DP. Idioms are discussed in section 4.

addressed in the current system. Rising is assumed, as stated at the end of the previous section. The perceived discontinuities are overcome by viewing the relevant catenae as attaching to words that dominate their governors. Risen catenae are indicated via dashed-dependency edges and the governor of a risen catena is marked with the *g* (= governor) subscript. Example (11b) illustrates *wh*-rising, (12b) scrambling rising, and (13b) extraposition rising. Our rising account of discontinuities is therefore assuming that projectivity violations never actually occur; our dependency structures are always projective.

A main aspect of rising like that in (11)–(13) is that the risen catena attaches to a word that dominates its governor. Risen catena is defined as follows:

Risen catena: A catena that is not immediately dominated by its governor

Given this definition, the central limitation on rising is expressed as follows:

Rising Principle: It must be the case that either (i) the head of a risen catena dominates the governor of that catena, or (ii) the risen catena itself dominates its governor.

One sees that in each of (11b), (12b), and (13b), part (i) of this definition is satisfied—that is, the head of the risen catena dominates that catena's governor. The head of *which explanation* in (11b) is *did*, whereby *did* dominates *produce*, the governor of *which explanation*. The head of *sich* 'self' in (12b) is *hat* 'has', whereby *hat* dominates *verletzt* 'injured', the governor of *sich*. And the head of *with long red hair* in (13b) is *arrived*, whereby *arrived* dominates *man*, the governor of *with long red hair*.

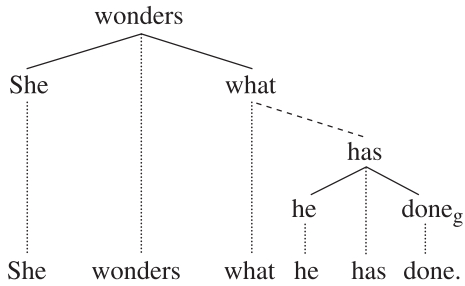
The concept of rising illustrated in (11)–(13) and expressed with the Rising Principle has many precedents in the dependency-grammar literature, although the terminology varies (see, e.g., Duchier & Debusmann 2001; Bröker 2000, 2003:294; Eroms & Heringer 2003:26). Although there are certainly differences between the accounts of these linguists, the underlying idea is the same. This idea is that a flattening of structure occurs in order to overcome the discontinuity.

2.4 Type 1 versus Type 2 Rising

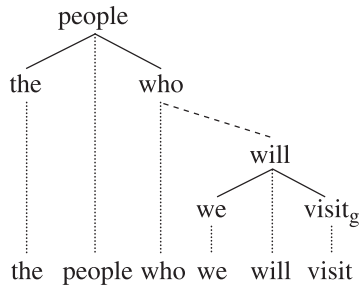
The rising illustrated in (11b), (12b), and (13b) has the head of the risen catena dominating the governor of that catena. In this regard, part (i) of the Rising Principle is satisfied. Part (i) of the Rising Principle identifies *Type 1 rising*. The most distinctive trait of Type 1 rising is that the risen catena is a constituent. Each of the risen catenae in (11b), (12b), and (13b) is a constituent (as defined in section 2.2).

Part (ii) of the Rising Principle identifies *Type 2 rising*. Type 2 rising obtains when the risen catena itself dominates its governor. This occurs in indirect questions and relative clauses. The root of the indirect question or relative clause, which is usually the *wh*-element or relative pronoun, is (the root of) the risen catena:

(14)



(15)



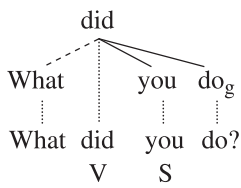
The risen catena is *what* in (14) and *who* in (15). These catenae dominate their governors—that is, *what* dominates *done* and *who* dominates *visit*. Part (ii) of the Rising Principle is therefore satisfied. The dashed dependency edges in (14) and (15) are consistent in that they continue to indicate a constituent the head of which is not its governor. Thus in these cases, the *wh*-element is not the governor of the clause that it immediately dominates.[9] The *g* subscripts are also consistent insofar as they continue to mark words that are not the head of a catena that they govern.

The unusual and perhaps confusing aspect of Type 2 rising is that the risen catena is *not* a constituent and therefore appears to have two governors. In (14), the risen catena *what* appears to have both *wonders* and *done* as its governor, and in (15), the risen catena *who* appears to have both *people* and *visit* as its governor. Appearances are misleading, however. The governor of the risen catena in each case is marked by the *g* subscript. *Wonders* in (14) is the governor of the entire interrogative clause (not just of *what*), and *people* in (15) is the governor of the entire relative clause (not just of *who*).

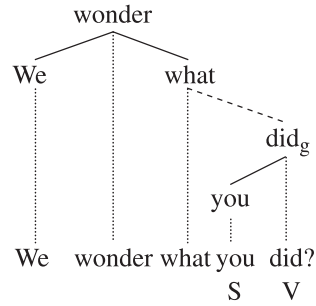
A primary source of support for the distinction between Type 1 and Type 2 rising occurs across interrogative matrix and subordinate clauses. *Wh*-fronting in matrix clauses results in VS (= aux + subject) order, whereas *wh*-fronting in subordinate clauses results consistently in SV order.

[9] That the *wh*-element does not license the clause that it immediately dominates in cases like (14) and (15) can be seen in the fact that *wh*-elements in direct questions do not license such clauses, as in **What he said did she hear?*.

(16) a.



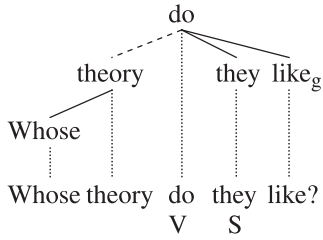
b.



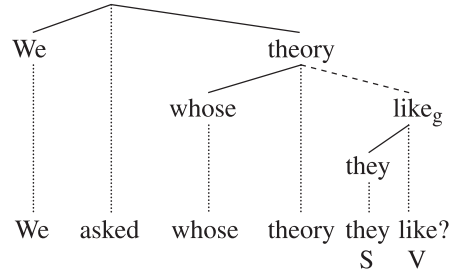
By viewing the *wh*-element as the root of the interrogative clause in (16b), we have a principled means of addressing the contrast between VS and SV order. When VS order obtains as in (16a), the *wh*-element is a dependent of the finite auxiliary, which means Type 1 rising is present. But when SV order obtains as in (16b), the *wh*-element immediately dominates the finite verb, which means Type 2 rising has occurred.

When a *wh*-element or relative pronoun pied-pipes material, the root of the pied-piped catena can be the root of a subordinate clause, as in the trees in (17b) and (18b).

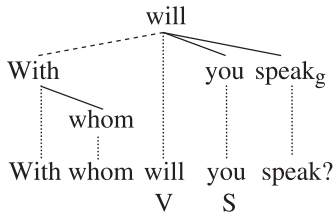
(17) a.



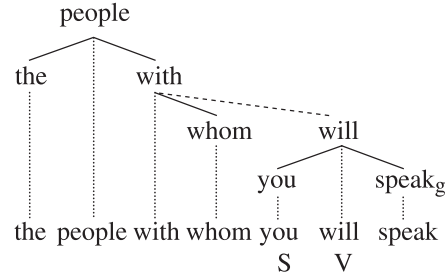
b.



(18) a.



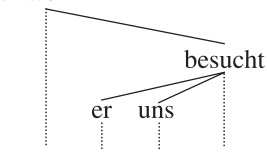
b.



The *wh*-elements pied-pipe their governors/heads. The two together, the *wh*-element and its governor/head, are the risen catena in each case. The examples illustrate again the distinction between VS and SV order. When VS order obtains, Type 1 rising is present; when SV order obtains, Type 2 rising has occurred.

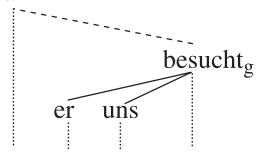
A second and similar observation that supports the distinction between Type 1 and Type 2 rising comes from German. Subordinators in German force verb-final order. Similarly, *wh*-elements (in subordinate interrogative clauses) and relative pronouns necessitate verb-final order. Observe the parallelism in word order across the following (a) and (b) examples:

(19) a. wenn



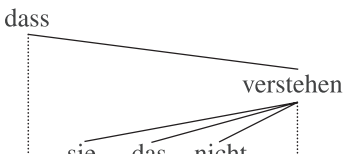
...wenn er uns besucht
when he us visits
'when he visits us'

b. wann



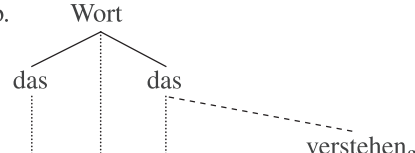
(Wir wissen,) wann er uns besucht.
we know when he is visits
'We know when he is visiting is.'

(20) a.



dass sie das nicht verstehen
that they that not understand
'that they don't understand that'

b.



das Wort, das sie nicht verstehen
the word that they not understand
'the word that they don't understand'

The subordinator *wenn* in (19a) forces verb-final order, as opposed to the standard V2 order of matrix clauses in German, as in **wenn er besucht uns*. The same verb-final order occurs with the interrogative adverb *wann* in (19b). Similar remarks apply to the subordinator *dass* 'that' and the relative pronoun *das* 'that' in (20). The parallelism suggests therefore that the subordinators *wenn* and *dass*, the interrogative adverb *wenn*, and the relative pronoun *das* have the same impact on word order. Given that dependency grammars and constituency grammars alike assume that subordinators like *wenn* and *dass* are the roots/heads of the clauses they introduce and *wh*-elements and relative pronouns introducing subordinate clauses force the same word order as subordinators, we can assume that these word classes occupy the same position in the structure. Type 2 rising allows us to make this assumption. If Type 1 rising were all that the theory had at its disposal, the parallelism could not be accommodated. Groß & Osborne (2009) provide a more detailed discussion of these and other aspects of rising.

A final note about rising is warranted. The term *rising* suggests a derivational view of syntax. The current dependency grammar is, however, decidedly representational. Thus a risen catena should *not* be seen as having appeared as a dependent of its governor at some stage or point in a derivation below or beyond the surface, but

rather the term *rising* is understood metaphorically. It is intended merely to denote a constellation in which a word fails to immediately dominate a catena that it governs.

3. Catenae and Constituents

The discussion now returns to the catena. The catena as defined and briefly illustrated in section 2.2 is the fundamental unit of syntax in dependency grammar and presumably in grammar in general.

Claim 1: The catena is the fundamental unit of syntax, not the constituent.

This claim expresses the central message of this paper—namely, that the catena should replace the constituent as the fundamental unit of syntax. The other two claims mentioned in the introduction, claims 2 and 3, are in some sense derived from claim 1. The rest of this paper strives to establish a clear understanding of catenae and to demonstrate the utility of the concept. In so doing, claim 1 is defended.

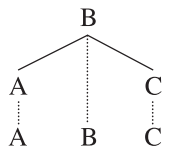
In graph-theoretic terms, a catena is any *tree* or *any subtree* (complete or partial) of a tree, whereby the number of complete and partial subtrees for a given tree is always limited. To get a sense of what the catena can accomplish, it is necessary to first acknowledge some aspects of word combinations in dependency (and constituency) trees. The dependency-grammar catena is a much more inclusive unit of syntax than the constituency-grammar constituent. By ''more inclusive,'' we mean that many more word combinations in a given structure qualify as catenae than as constituents. Recall that the constituent is a subtype of catena.

The number of all possible distinct word combinations for a given tree is calculated using the following formula:

(21) $2^n - 1$ where n = the number of words

The following trees illustrate. The capital letters represent words.

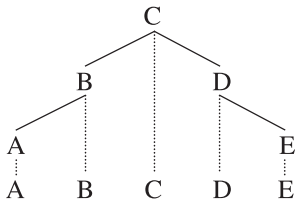
(22)



$2^3 - 1 = 7$

Distinct combinations: A, B, C, AB, BC, AC, ABC

(23)



$2^5 - 1 = 31$

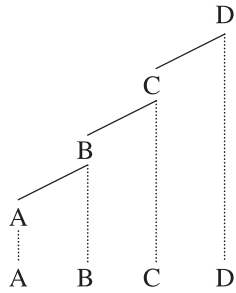
Distinct combinations: A, B, C, D, E, AB, AC, AD, AE, BC, BD, BE, CD, CE, DE, ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE, ABCD, ABCE, ABDE, ACDE, BCDE, ABCDE

The number of distinct word combinations for each tree is calculated on the right using the formula. All of these combinations are then listed under this number.

The number of distinct catenae in a given tree is determined by counting them individually.[10] Once this number has been determined, the number of noncatena word combinations is easily calculated via subtraction. There are six distinct combinations that qualify as catenae in (22): A, B, C, AB, AC, and ABC. Therefore only one word combination (7 − 6 = 1) in (22) is a noncatena (namely, AC). There are 15 distinct combinations in (23) that qualify as catenae: A, B, C, D, E, AB, BC, CD, DE, ABC, BCD, CDE, ABCD, BCDE, and ABCDE. That means there are 16 noncatena combinations (31 − 15 = 16) in (25): AC, AD, AE, BD, BE, CE, ABD, ABE, ACD, ACE, ADE, BCE, BDE, ABCE, ABDE, and ACDE.

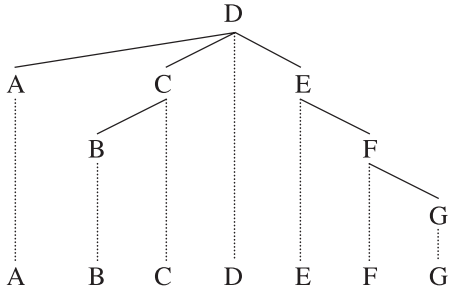
These numbers can be expressed as ratios, whereby the first number in the ratio is the number of distinct catenae and the second number in the ratio is the number of distinct noncatenae. Thus the ratio for tree (22) is 6:1, and the ratio for tree (23) is 15:16. Such ratios shall be called *catena ratios*. The following trees illustrate catena ratios further:

(24)



Catena ratio: 10:5

(25)



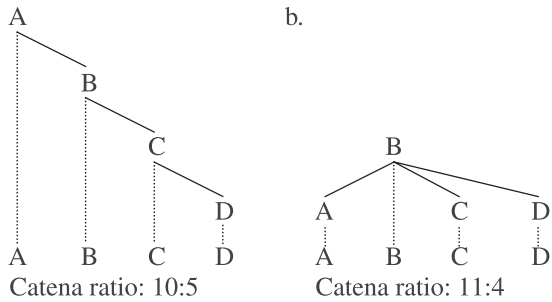
Catena ratio: 33:94

[10] Calculating the number of distinct catenae is a complex polynomial problem. The discussion here avoids this problem, opting instead to simply count the catenae individually.

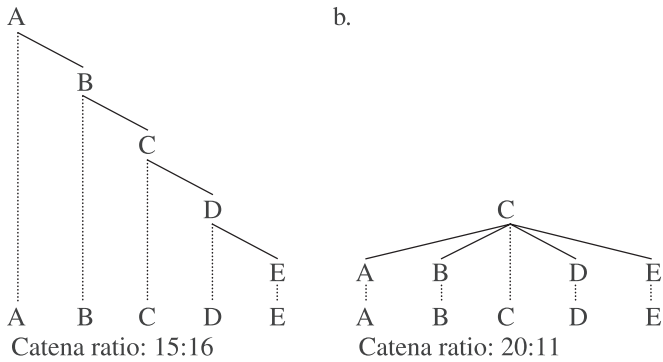
Comparing these ratios, one sees that the percentage of noncatenae word combinations increases as the number of words increases.

An interesting aspect of such ratios is that flatter structures contain more catenae than more layered structures. This point is illustrated with the following trees:

(26)



(27)

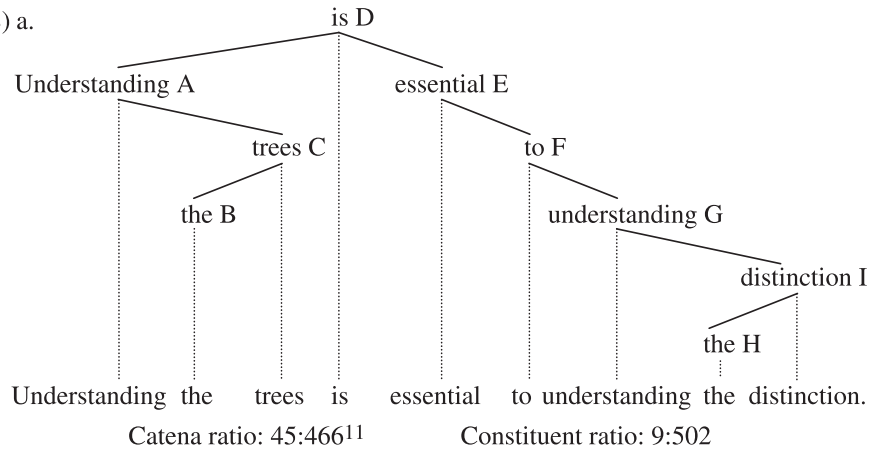


Each tree of a pair contains the same number of nodes. The important difference between the trees lies with the number of levels. The more layered trees on the left contain fewer catenae than the corresponding flatter trees on the right.

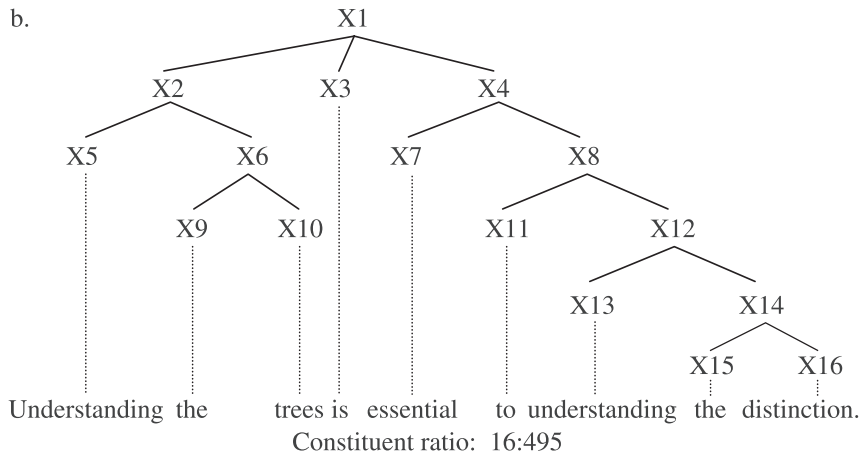
The dependency-grammar catena is a more inclusive unit of syntax than the constituency-grammar constituent. The following trees illustrate this point:[11]

[11] All 45 catenae in (28a) are listed here: A, B, C, D, E, F, G, H, I, AC, AD, BC, DE, EF, FG, GI, HI, ABC, ACD, ADE, DEF, EFG, FGI, GHI, ABCD, ACDE, ADEF, DEFG, EFGI, FGHI, ABCDE, ACDEF, ADEFG, DEFGI, EFGHI, ABCDEF, ACDEFG, ADEFGI, DEFGHI, ABCDEFG, ACDEFGI, ADEFGHI, ABCDEFGI, ACDEFGHI, and ABCDEFGHI.

(28) a.



b.



The catena ratio for the dependency tree (28a) is 45:466. We have also included the constituent ratio for (28a)—that is, the ratio of constituent to nonconstituent word combinations. Because there are nine words, there are nine constituents. In contrast, the number of constituents in the constituency tree (28b) is 16. The important thing to acknowledge about these numbers is that there are usually significantly more catenae in a given structure than there are constituency-grammar constituents. In this regard, the catena is indeed a much more inclusive unit of syntax than the constituent.

4. Idiomatic Meaning

The catena concept provides a direct connection to the semantics; semantic units are stored as catenae and often appear as catenae in the syntax. Two simple

examples from the introduction provide a preliminary illustration of this point:

(29) She *has lost* her keys.

(30) Fred *took* us *on*.

In isolation, the verb *has* means 'possesses, owns', yet when it combines with a participle such as *lost* in (29), its meaning shifts drastically. It becomes an auxiliary of aspect devoid of lexical content. The auxiliary and participle combine to form a semantic unit—namely, a predicate that expresses perfective aspect. This predicate, *has lost*, does not, however, form a constituent in surface syntax (in most theories).

The same situation is true of the phrasal verb *took . . . on* in (30). The verb *took* and its particle *on* clearly combine to form a semantic unit, 'take on', to the exclusion of the object pronoun *us*. Yet *took* and *on* do not form a constituent in surface syntax such that this constituent excludes *us*. This noncorrespondence between semantic units and constituents is not a problem if one takes the catena as the fundamental unit of syntax, since the word combinations *has lost* and *took . . . on* are catenae.

The insight concerning (29) and (30) can be extended to idiomatic meaning of all types, as the next two subsections make clear.

4.1 Nonconstituent Idioms

O'Grady (1998) demonstrates that the words constituting idioms are always catenae in the lexicon (see also Hyvärinen 2003:749–750). Osborne (2005) builds on O'Grady's account of idioms. He observes that when a predicate includes one word of a multiple word idiom, then it includes all the words of that idiom. What these analyses of idioms make evident is that the catena is the relevant unit of syntax for a theory of idioms, not the constituent. This insight is expressed as claim 2:

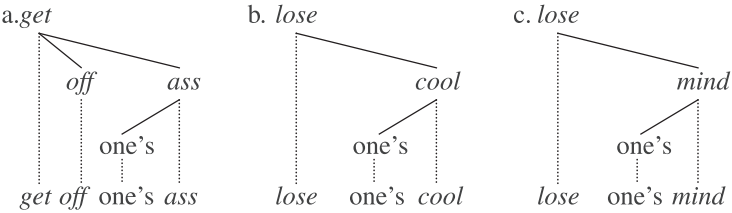
Claim 2: All idioms are stored as catenae, but not all idioms are stored as constituents.

The discussion in this section establishes the validity of this claim.

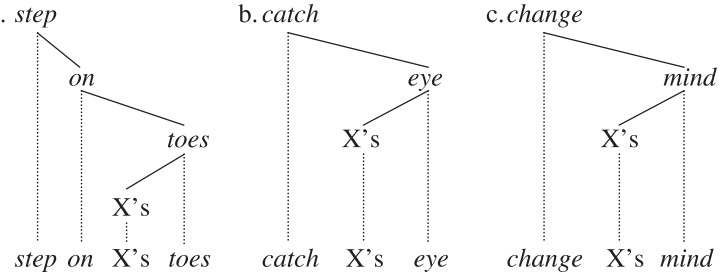
The literature (e.g., Nunberg, Sag & Wasow 1994; Horn 2003) acknowledges various types of idioms. Some idioms are relatively free (e.g., *keep tabs on*) insofar as they undergo various syntactic processes, whereas other idioms (e.g., *kick the bucket*) cannot undergo these processes and are therefore fixed. All idioms are created equal, however, insofar as they are all stored as catenae. In other words, they are catenae in their lexical entries. Crucially though, there are very many idioms that are nonconstituents in their lexical entries.

Many idioms include a verb and a noun but exclude the noun's determiner. The words of the idioms are in italics:

(31)



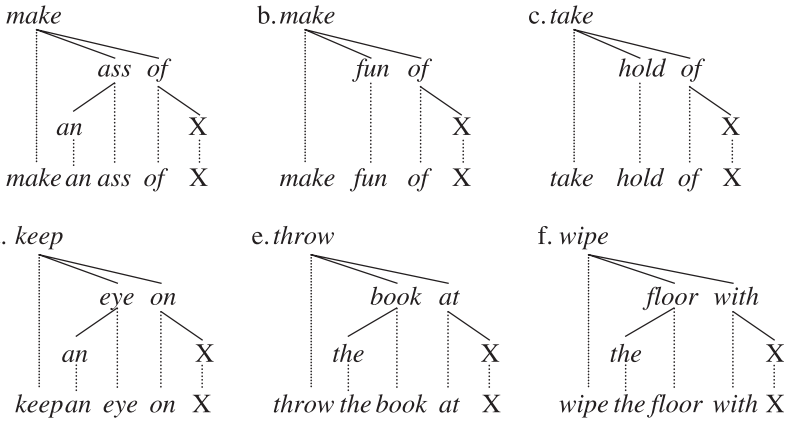
(32)



Each time the words of the idiom form a catena but not a constituent. For instance, *lose . . . cool* in (31b) is a catena but not a constituent, and *step on . . . toes* in (32a) is a catena but not a constituent. Consider what a derivational constituent-based approach would have to do to account for such idioms. It would have to posit, for instance, that *step*, *on*, and *toes* in (32a) form a constituent (to the exclusion of the possessor) at some point in the derivation or at some level of representation below or beyond the surface. Such an account now seems implausible in light of the fact that these three words straightforwardly form a catena as shown.

There are also many idioms that include a verb and a preposition but that exclude the object of the preposition:

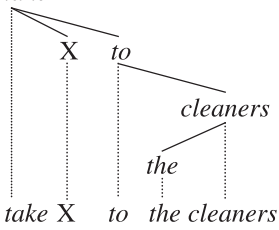
(33)



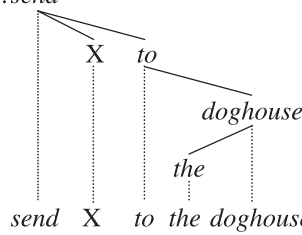
Each time the idiom excludes the argument marked by X, and each time the words of the idiom form a catena but not a constituent. For example, *make an ass of* in (33a) is a catena but not a constituent, and *wipe the floor with* in (33f) is a catena but not a constituent.

Many idioms include an object NP or PP but exclude the (other) object NP:

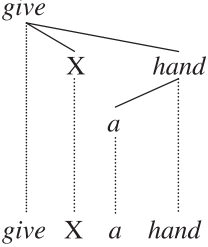
(34) a. *take*



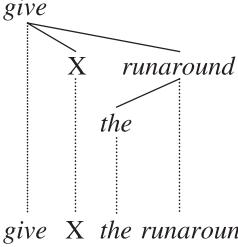
b. *send*



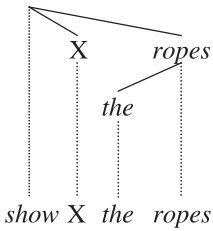
c. *give*



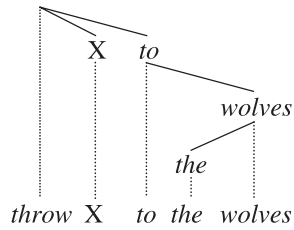
d. *give*



e. *show*



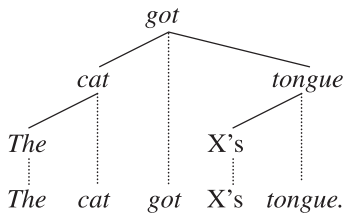
f. *throw*



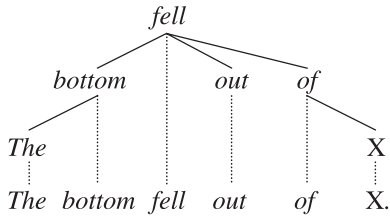
The italicized words in each case form a catena but not a constituent. For instance, *send . . . to the doghouse* in (34b) is a catena but not a constituent, and *throw . . . to the wolves* in (34f) is a catena but not a constituent.

Some idioms even include the subject NP but exclude some other expression:

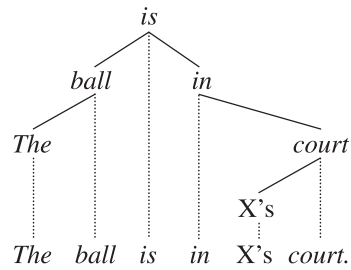
(35)



(36)



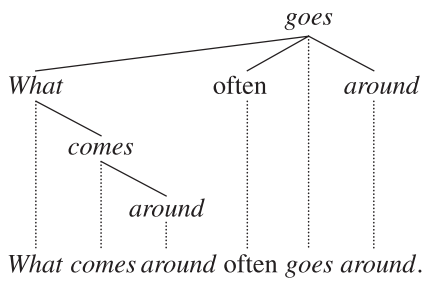
(37)



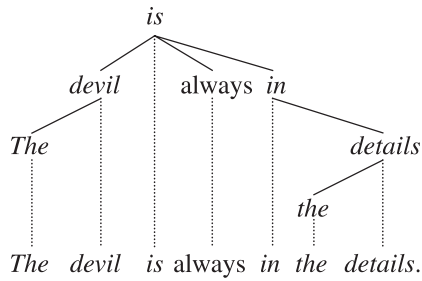
These idioms are unusual insofar as idioms that include the subject at the same time that they exclude some other constituent lower in the hierarchy are rare. The key point to acknowledge about all of (31)–(37) is that although the words forming the idioms are stored as catenae, they cannot be viewed as being stored as constituents. The number of idioms that fail to qualify as constituents is very large. Spending some time with an idiom dictionary verifies that this is so.

Sayings can be viewed as idioms that encompass the entire sentence. Adjuncts can modify sayings, but when they do, the words of the saying form a catena to the exclusion of the adjunct.

(38)



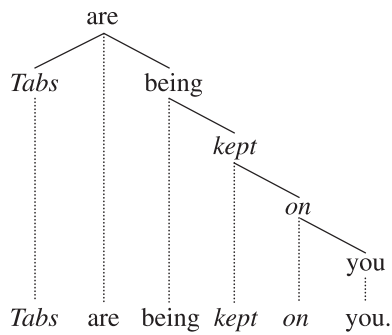
(39)



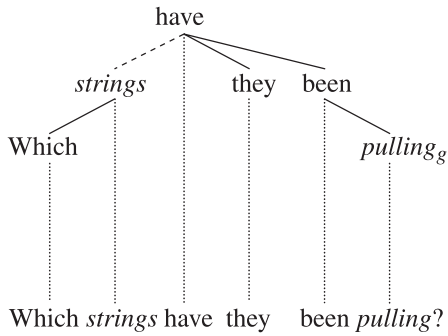
The adverbs *often* in (38) and *always* in (39) are outside of the saying in each case. Accordingly, the words of the saying form a catena to the exclusion of the adverb. It is difficult to see how such data could be analyzed in terms of constituents.

An important limitation on this account of idioms must be acknowledged. The qualifier ''are stored as'' (as opposed to just ''are'') in claim 2 points to the fact that the words of idioms can be broken up by syntactic processes, for instance:

(40)



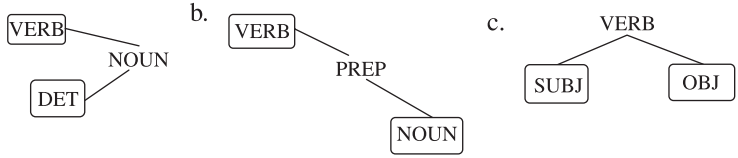
(41)



The words of the idioms (in italics) in (40) and (41) do not form catenae. Syntactic processes—for instance those that express aspect and voice and result in long-distance dependencies—can break up idiom catenae. In light of such data, the qualifier ''stored as'' in claim 2 is necessary.

The analysis of idioms just given makes firm predictions about the nature of idioms and the word combinations that can form idioms. It explains Sportiche's (2005:79) observation that certain word combinations never form idioms.

(42) a. b. c.



One does not, for instance, find verb–determiner idioms to the exclusion of the noun in a constellation like (42a), nor does one encounter verb–noun idioms to the exclusion of the preposition in a constellation like (42b), and subject–object idioms to the exclusion of the lexical verb in a constellation like (42c) also never occur. One can search idiom dictionaries in vain; such idioms do not exist.

Sportiche's (2005:79ff.) discussion of the words that form idioms should be, as a point of comparison, insightful. Sportiche observes that there is indeed something

unique about the words that can form idioms. This uniqueness is manifest in two easily observable facts: first, the word combinations that form idioms are not arbitrary, and second, a (constituent-based) analysis of these combinations is challenged in more ways than one. Sportiche (2005:80–81) then offers the following observation about the words that form idioms:

> If Y is the highest head (for c-command) of the idiom and W its lowest, all intermediate heads (e.g., Z) must be part of the idiom. Specifiers of and adjuncts to the heads can be but do not have to be part of the idiom: the sequence of idiomatic heads on the spine of the tree must be uninterrupted.

The similarity of this observation with the catena concept should be apparent. Sportiche's insight can be expressed straightforwardly using catenae.

Finally, the vP-internal-subject hypothesis (Larson 1988, Koopman & Sportiche 1991) should be mentioned. A primary motivator for this hypothesis is that those idioms that include the subject can be broken up by auxiliaries, as in *The shit hit the fan*, *The shit will hit the fan*, *The shit has been hitting the fan*, *The shit will have been hitting the fan*, and so on (see, e.g., Koopman & Sportiche 1991:224–225). These data are taken to indicate at least two things: first, the subject is generated first in the specifier position of vP, and second, idioms that include the subject exclude tense. Each of these points is now examined in light of the current catena-based analysis.

If the subject of the idiom appears first in Spec,vP, it can be seen as forming a constituent there with the rest of the idiom. From its position in Spec,vP, the subject then moves up the structure to a higher position (e.g., to Spec,I/TP), which explains the intervention of auxiliaries between the words of the idiom. In this manner, a constituent-based analysis can be assumed for idioms that include the subject. That is, the words of the idiom form a constituent (below the surface) after all, contrary to appearances.

The data produced above cast doubt on the validity and utility of this analysis and thus on the vP-internal-subject hypothesis in general. Numerous idioms that exclude the subject (e.g., *wipe the floor with X*, *throw X to the wolves*, *make fun of X*) can in no way be viewed as forming constituents below the surface. Furthermore, at least some idioms that do include the subject also cannot be viewed as forming constituents below the surface (e.g., *The cat got X's tongue*, *The bottom fell out of X*, *The ball is in X's court*). What these data demonstrate, then, is that the inclination to see idioms forming constituents at some point in the derivation simply cannot be maintained. This aspect of the vP-internal-subject hypothesis is without merit.

The second point—namely, that those idioms that include the subject always exclude tense—does support the vP-internal-subject hypothesis, however. We agree with the basic insight insofar as idioms that include the subject exclude tense. The insight does not, however, challenge our analysis. Our idioms are stored as catenae without tense, as in *The shit hit the fan*, *All hell break loose*, and *The devil be in the details*. In this regard then, there is some overlap between the vP-internal-subject hypothesis and our understanding of the mental lexicon. The former sees the words of idioms entering the derivation low down in the structure where tense is absent, whereas the latter sees idioms being stored without tense.

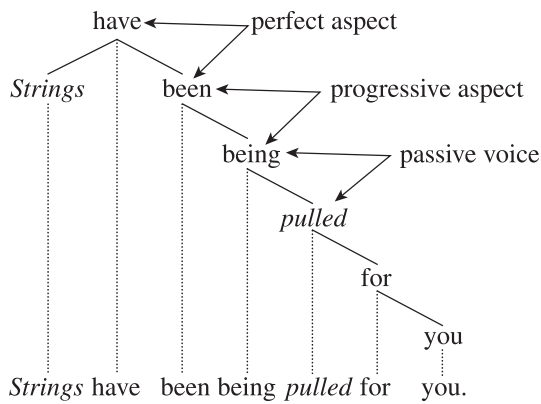
4.2 Overlapping Constructs

Examples (40) and (41) illustrated that the words of idioms can be broken up by syntactic processes. An anonymous reviewer asks about the nature of these syntactic processes and an editor requests more discussion of the exclusion of tense from those idioms that include the subject. This section briefly sketches the current theory's account of these areas.

Our dependency-based theory acknowledges constructions in the sense of Construction Grammar (CxG), following the tradition of linguists such as Lakoff (1987), Fillmore, Kay & O'Connor (1988), Kay & Fillmore (1999), and Goldberg (1995, 2006). We believe that most of the constructions discussed in the CxG literature are stored as catenae on the lexicon–syntax continuum. In other words, constructions are catenae. One must acknowledge in this regard, however, the distinction between constructions and constructs (type vs. token). Constructions are abstract entities that appear as catenae on the syntax–lexicon continuum, whereas constructs are the concrete manifestations of constructions in the syntax of actual utterances. Constructions are always catenae, but the corresponding constructs may or may not appear as catenae in actual syntax.

A given construct fails to qualify as a catena in surface syntax when it is interrupted by or intersects with one or more other, more syntactic constructs. The idiom *pull strings* and periphrastic verb constructs of perfect and progressive aspect and passive voice are used to illustrate the point.

(43)



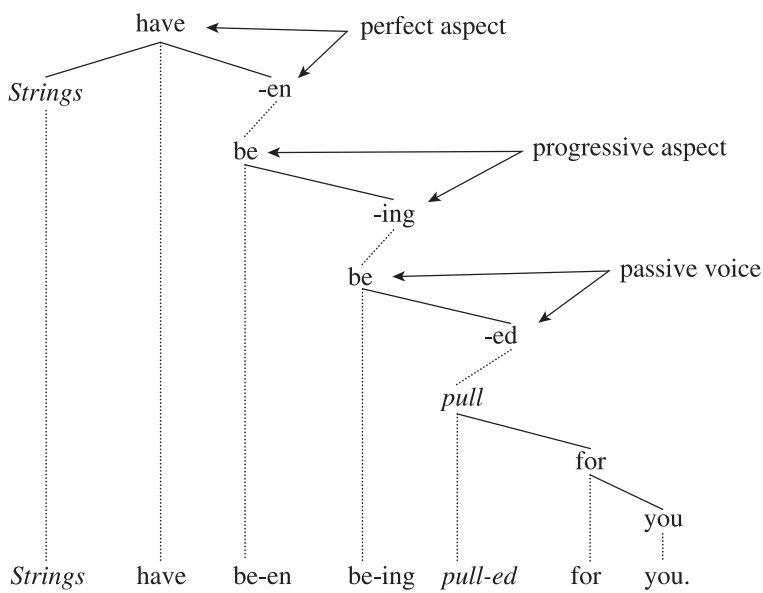
The words of the idiom are in italics. These words fail to form a catena because they are interrupted by the words of the perfect aspect construct (*have* and *been*) and the progressive aspect construct (*been* and *being*), and they partially overlap with the words of the passive voice construct (*being* and *pulled*). Note, however, that all four constructs together form a single catena to the exclusion of *for you*.

Example (43) illustrates what is meant in the previous section with the designation *syntactic processes*. This term denotes the interruption of a given construct by one or more, more syntactic constructs. In the case of (43), the given construct is the idiom *strings . . . pulled* and the more syntactic constructs are *have been*, *been being*, and

being pulled. Thus idiom constructs, which are stored as catenae on the lexicon–syntax continuum (i.e., they are constructions), can be broken up in the actual syntax by virtue of the fact that they co-occur with more syntactic constructs. When this occurs, the relevant constructs together always form a single greater catena. In this case, this greater catena is *Strings have been being pulled*. Note that this account acknowledges surface syntax only. The constructions exist as abstract entities on the lexicon–syntax continuum, but the corresponding constructs occur and co-occur in actual syntax. Syntactic phenomena that are addressed in terms of movement in derivational theories are addressed in the current system in terms of co-occurring constructs.

The current theory's account of co-occurring constructs can be much more exact than tree (43) suggests. Our system acknowledges both interword (= between words) and intraword (= within words) dependencies. The dependencies discussed so far have all been interword ones. Intraword dependencies, in contrast, are dependencies between the morphs that constitute words (see Groß 2010). In other words, our system extends dependencies into the morphosyntax. Suffixes, because they impact the distribution of their host words, are assumed to be the roots of their words. Dotted dependency edges mark the intraword dependencies. Example (43) receives the following analysis:

(44)



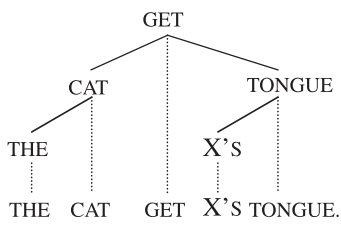
The tree is consistent with respect to both the interword-dependency edges and the projection edges. Just the intraword-dependency edges (dotted) have been added. This analysis grants the morphs *-en*, *-ing*, and *-ed* nodes in the structure. Our DG analysis in (62) is reminiscent of the structures resulting from Brody's (2000)

Telescope Principle in Mirror Theory. In our understanding, Mirror Theory is inherently dependency-based but lacks the catena concept.

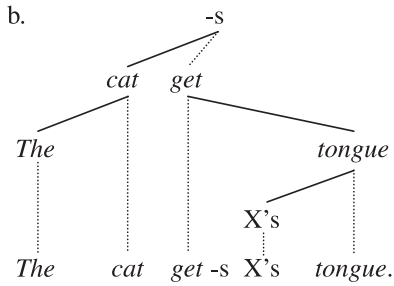
The advantage to the representation (44) is that the periphrastic verb constructs can now be clearly delineated. The nodes *have* + *-en* form the perfect-aspect construct; the nodes *be* + *-ing* form the progressive-aspect construct, and the nodes *be* + *-ed* form the passive-voice construct. Each of these constructs is a catena. The term *morph catenae* is employed to denote such catenae. Morph catenae are the means by which the current system addresses the auxiliary system. Constituency-based theories, in contrast, traditionally address such phenomena in terms of some sort of rightward movement, such as affix hopping (Chomsky 1957) or lowering and/or local dislocation (Halle & Marantz 1993, Embick & Noyer 2001).

Morph catenae of the sort illustrated in (44) are important for the analysis of tense in idioms. Idioms that include the subject NP are constructions just like any other idioms. These constructions lack tense. When they appear in actual syntax, however, the subject will often fail to form a catena with the other words of the idiom because at the very least, a tense/person construct intervenes. The point is illustrated with the following trees:

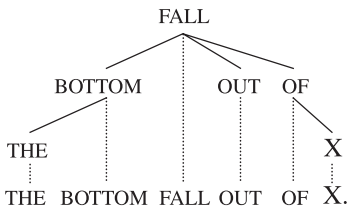
(45) a.



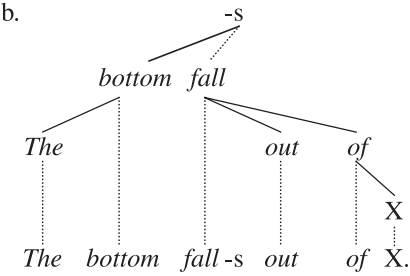
b.



(46) a.



b.



The small caps on the left indicate constructions (not constructs). The verb of these constructions lack morphology of tense and person. The trees on the right, in contrast, show constructs that correspond to these constructions. The words of each idiom are in italics. Note that the suffix *-s* indicating present tense and third-person singular is *not* part of the idiom in each case. The words of each idiom fail to form a catena precisely because the suffix *-s* of tense and person—this suffix is a

construct—intervenes. The two constructs together, however, form a single catena. Therefore we see again that the words of an idiom can be broken up by other constructs in actual syntax, in this case by the tense/person suffix *-s*.

This section has sketched the current theory's understanding of the issues concerning the ''syntactic processes'' that break up idiom catenae. Furthermore, the manner in which tense can be separated from those idioms that include the subject has been outlined in terms of morph catenae. Of course a principled account of these issues requires much more discussion than we have provided here. The focus of this paper does not lie with morph catenae, however, but rather with word catenae.

5. Ellipsis

The catena is the key unit of syntax for many ellipsis mechanisms.

Claim 3: The elided material of many ellipsis mechanisms (answer fragments, gapping, stripping, VP ellipsis, pseudogapping, sluicing, comparative deletion) is always a catena, but not always a constituent.

The following sections establish the validity of this claim by examining the named ellipsis mechanisms. It should be noted from the outset that claim 3 is a necessary condition on these ellipsis mechanisms but not a sufficient one.

5.1 Movement First, Ellipsis Second

Before examining the seven named ellipsis mechanisms directly, we briefly consider the approach to ellipsis common to many constituency-based theories. This approach assumes movement first and ellipsis second. That is, the remnants move out of the relevant constituent first before ellipsis occurs. In this manner, ellipsis can be viewed as operating on constituents only. Examples (47) and (48) illustrate this type of approach:

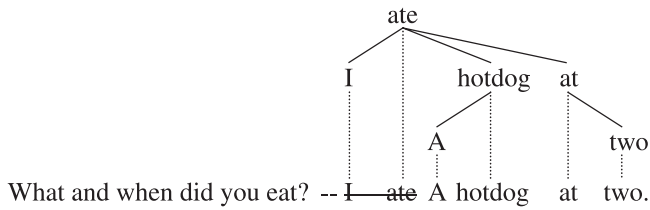
(47) What did Susan eat? Beans$_i$ ([$_{TP}$ Susan ate t_i]).

(48) Susan should [$_{VP}$ eat beans], and
Fred should rice$_i$ [$_{VP}$ t_i].

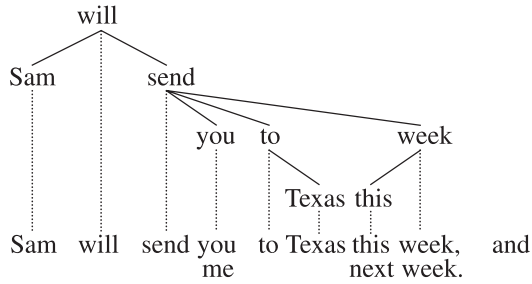
Merchant (2004) addresses answer fragments like the one in (47) in terms of topicalization. The answer fragment *Beans* is fronted in such a manner that the entire TP can then be elided (like in cases of sluicing). This movement ensures that ellipsis, contrary to first impression, is operating on constituents after all. Accounts of pseudogapping are often similar (e.g., Kuno 1981, Jayaseelan 1990, Lasnik 1999). The object *rice* in (47) moves to a higher projection (i.e., either Spec,Agr$_O$P or to the phrasal edge of vP) so that ellipsis can then elide the empty VP.

A problem facing this sort of approach to ellipsis is apparent when the number of remnants that must move out of the relevant constituent exceeds one. The following instances of an answer fragment and gapping illustrate the problem:

(49)



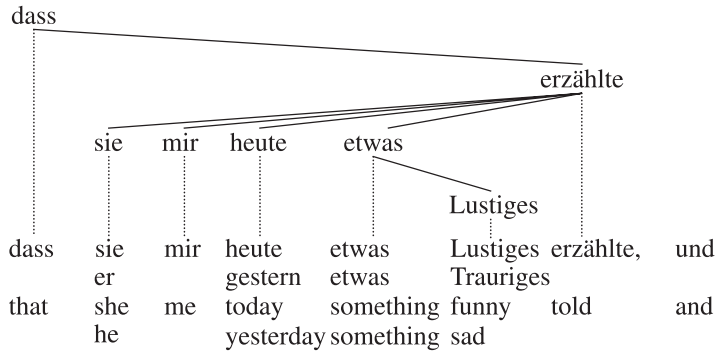
(50)



Both remnants in the answer fragment in (49) must be fronted so that the TP can be elided. Similarly, both remnants in the instance of gapping in (50) must exit the relevant constituent so that this constituent can be elided. Confronted with the necessity of multiple movements in such cases, the concern arises as to whether the movement account is economical. This concern grows when one acknowledges that the catena-based analysis does not need movement. The elided material in each case is a catena; the elided *I ate* in (49) is a catena, and the elided *Sam will send . . . to Texas* in (50) is a catena.

The problem is perhaps more evident in other languages. Consider the following example from German:

(51)



'that she told me something funny today, and he something sad yesterday'

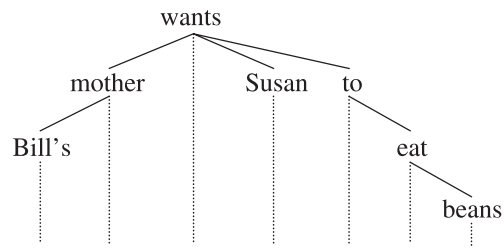
If one assumes a finite-VP constituent for German, then two remnants must be assumed to exit this VP so that ellipsis can occur, and if one assumes a flat structure (lacking a finite VP), then all three remnants must exit the relevant constituent, whatever that constituent is deemed to be. A further problem is that the types of movement involved could hardly be the same movement mechanisms as for gapping in English. Topicalization does not so obviously exist in subordinate clauses in German (although scrambling of course does), and extraposition in German always has the extraposed constituent(s) appearing to the right of the clause-final verb(s), not to the left of it as in (69). These problems do not exist for the catena-based approach. The elided material *mir . . . erzählte* is straightforwardly a catena.

We do not dispute that movement analyses of data like (49)–(51) already exist or could be devised. What we dispute is that such analyses are as parsimonious as our account in terms of catenae. The elided material in (49)–(51) is a catena on the surface; there is no necessity to assume movement in order to address such data.

5.2 Answer Fragments

The elided material of standard answer fragments is always a catena. This material usually fails to qualify as a constituent, however:

(52)



a.	Whose mother wants Susan to eat beans?	Bill's ~~mother wants Susan to eat beans~~.
b.	Who wants Susan to eat beans?	Bill's mother ~~wants Susan to eat beans~~.
c.	Who does Bill's mother want to eat beans?	~~Bill's mother wants~~ Susan ~~to eat beans~~.
d.	What does Bill's mother want Susan to do?	~~Bill's mother wants Susan to~~ Eat beans.
e.	What does Bill want Susan to eat?	~~Bill's mother wants Susan to eat~~ Beans.

The elided material of each answer fragment is struck out. Each time these words form a catena. For instance, *Bill's mother wants . . . to eat beans* in (52c) is a catena, but certainly not a constituent, and *Bill's mother wants Susan to* in (52d) is a catena, but certainly not a constituent. Note further that each answer fragment in (52a–e) is a dependency-grammar constituent.

When the answer fragment is not a dependency-grammar constituent, the elided material is not a catena and the result is robustly unacceptable:

(52) f. Which relative of Bill's wants Susan to eat beans? *~~Bill's~~ Mother ~~wants Susan to eat beans~~.
g. How does Bill's mother view Susan eating beans? *~~Bill's mother~~ Wants ~~Susan to eat beans~~.
h. What does Bill's mother want Susan to do with beans? *~~Bill's mother wants Susan to~~ Eat ~~beans~~.
i. Who does Bill's mother want to eat beans? *~~Bill's mother~~ Wants Susan ~~to eat beans~~.
j. What does Bill's mother want Susan to do with beans? *~~Bill's mother~~ Wants Susan to eat ~~beans~~.

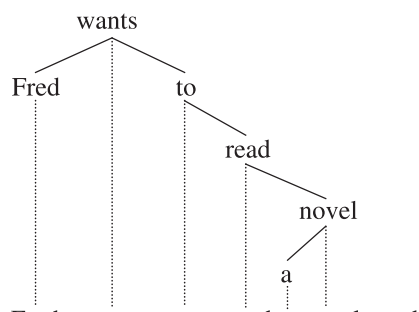
The answer fragment in each of (52f–j) fails because the elided material is not a catena. In (52f) for example, *Bill's . . . wants Susan to eat beans* is not a catena, and in (52j), *Bill's mother . . . beans* is also not a catena.

To be sure, there are numerous conceivable answer fragments that are disallowed despite the elided material qualifying as a catena, such as **Bill's . . . Susan*, **Bill's mother . . . Susan*, **Susan . . . eat beans*, and so on. Such data do not challenge claim 2 though, because claim 2 is a necessary condition on ellipsis, not a sufficient one. Such cases are accounted for largely by the additional requirement that answer fragments must be dependency-grammar constituents. The word combinations *Bill's . . . Susan*, *Bill's mother . . . Susan*, and *Susan . . . eat beans* are *not* dependency-grammar constituents.

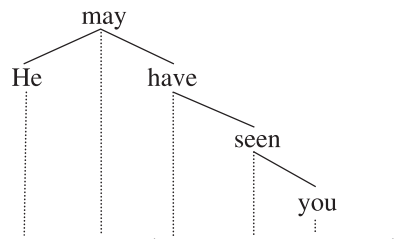
5.3 Gapping and Stripping

The words of the gaps of gapping must be catenae, whereby these catenae are often nonconstituents. If the words corresponding to the gap do not form a catena, gapping is impossible.

(53)



(54)

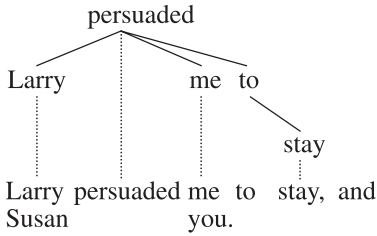


	He	may	have	seen	you,	and
a.	she				me.	
b.	*she			heard.		

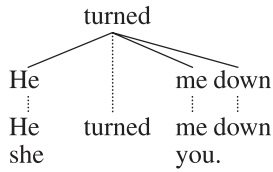
The gapped (a)-clauses, where the gap corresponds to a catena, are acceptable, whereas the gapped (b)-clauses, where the gap corresponds to a noncatena, are unacceptable. In other words, the word combination *wants to read* in (53) and *may have seen* in (54) are catenae and can be gapped, as illustrated in (53a) and (54a). The word combinations *wants . . . a novel* in (53) and *may have . . . you* in (54) are noncatenae and therefore cannot be gapped, as illustrated in (53b) and (54b). Note further that the gapped material in the (a)-clauses are not constituents.

In examples (53a) and (54a), the gap corresponds to a string of words. Gapping, however, allows the gapped words to be nonstrings. In such cases, these words still qualify as catenae:

(55)



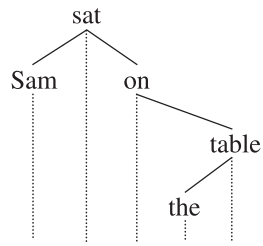
(56)



For similar examples, see Jackendoff 1971 and Neijt 1980:40. In these cases, the word combinations corresponding to the gaps are noncontiguous, that is, they are nonstrings. They are, however, catenae. In other words, *persuaded . . . to stay* in (55) and *turned . . . down* in (56) are nonstring (and nonconstituent) catenae.

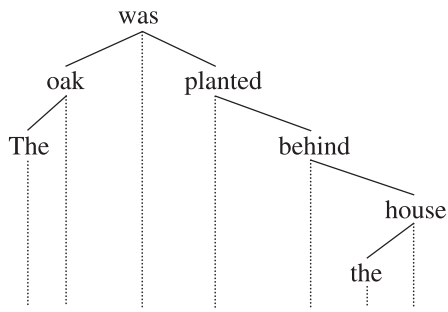
The following (b)-examples are similar to (55) and (56) insofar as the gap is noncontiguous. In contrast to (55) and (56) however, (57b) and (58b) are unacceptable because the words corresponding to the gaps do *not* form catenae:

(57)



Sam sat on the table, and
a. Susan under it.
b. *Susan under.

(58)



The oak was planted behind the house, and
a. the ash in front of it.
b. *the ash in front of.

In the (a)-clauses, the gap corresponds to just the verb(s), which is/are a catena, whereas in the (b)-clauses, the gap corresponds to the verb(s) plus the object of the PP, which together do not qualify as a catena.[12]

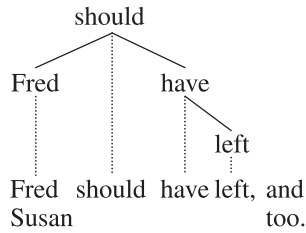
Stripping—a particular manifestation of gapping where only a single remnant appears in the gapped clause—also has the elided words corresponding to a catena.

[12] Note that the following examples are better than (53b) and (54b):

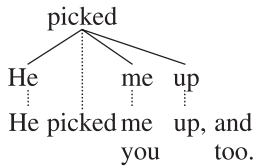
(i) [Fred sat on], and [Susan ___ under] the table.
(ii) [I am satisfied], but [you ___ dissatisfied] with the result.

These examples do not violate the catena requirement on gaps, given that they combine two sharing mechanisms—namely, gapping and so-called Right Node Raising. Only the finite verbs have been gapped. The material on the right periphery appears outside of the coordinate structure, as the brackets show.

(59)



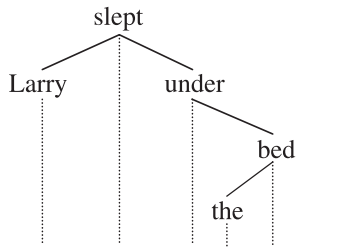
(60)



In each case, there is just a single remnant. In (59), this remnant is the subject, whereas in (60), it is the object. Despite the varying syntactic functions of the remnants, the elided words correspond to catenae. In (59), this catena is *should have left*, and in (60), it is *he picked . . . up*. Note further that the elided catenae *he picked . . . up* in (60) can in no way be understood as a constituent, regardless of the level of representation or point in the derivation that one chooses.

If the elided material of stripping fails to qualify as a catena, the result is unacceptable. This fact is illustrated in the following (b)-clauses:

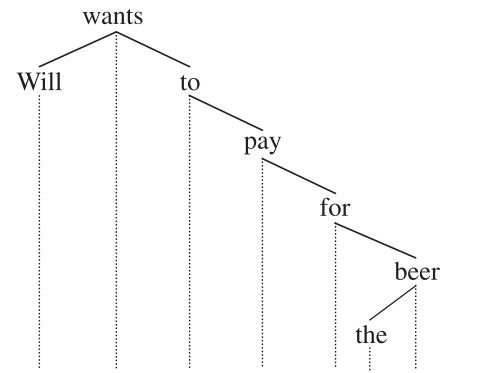
(61)



a. on it.

b. *on.

(62)

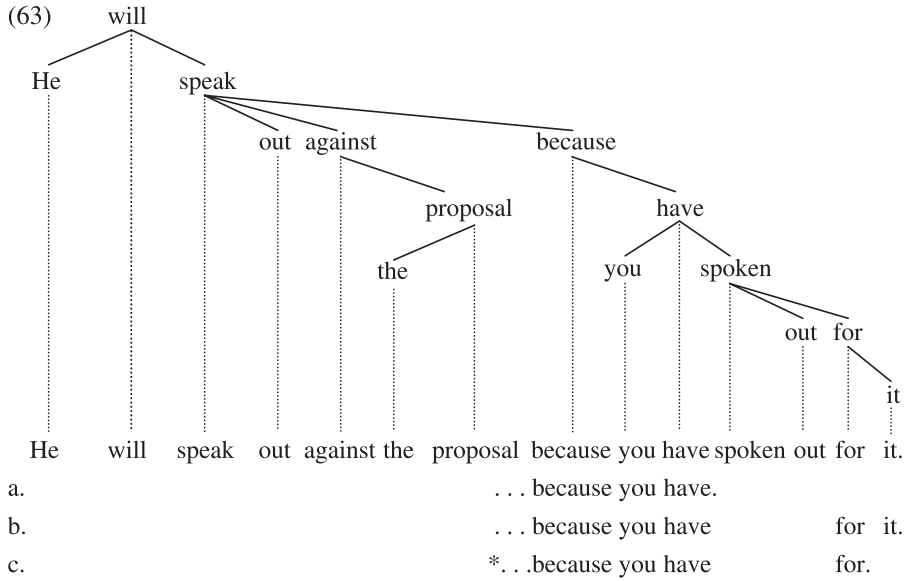


The gaps in the (a)-clauses correspond to nonconstituent catenae: *Larry slept* is a nonconstituent catena in (61) and *Will wants to* is a nonconstituent catena in (62). In contrast, the gaps in the (b)-clauses fail to qualify as catenae: *Larry slept . . . the bed* is not a catena in (61) and *Will wants to . . . the beer* is not a catena in (62).

5.4 VP Ellipsis and Pseudogapping

The elided material of standard instances of VP ellipsis is a constituent. The constituent is a subtype of catena, as established and emphasized in section 3.

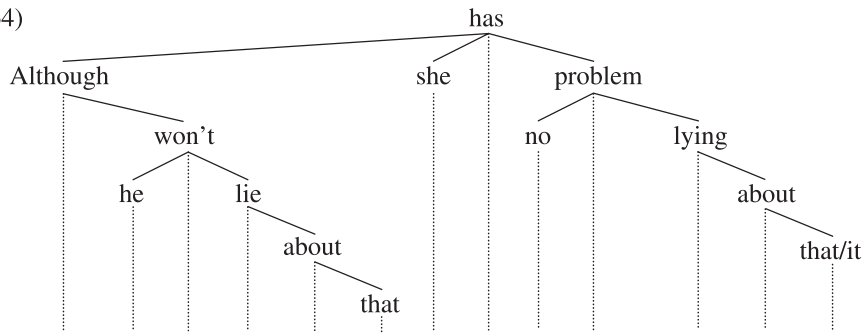
(63)



The elided material of the VP ellipsis in the subordinate clause in (63a)—namely, *spoken out for it*—is a catena. Example (63b) has the remnant *for it*, which makes it an instance of pseudogapping, and (63c) is disallowed because the elided *spoken out . . . it* is not a catena.[13]

Example (63) has the VP ellipsis following its antecedent. A well-known fact about VP ellipsis is that the ellipsis can precede its antecedent.

(64)



Although he won't lie about that, she has no problem lying about that/it.

a. Although he won't

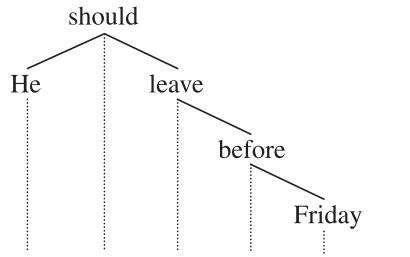
b.?? Although he won't about that,

c.* Although he won't about

The catena *lie about that* has been elided from the adjunct clause in (64a). When just the infinitive is elided, pseudogapping obtains and the result is at best strongly marginal, as shown in (64b). If, however, the elided material fails to qualify as a catena, the result is terrible, as shown in (64c).

Pseudogapping, which is a particular manifestation of VP ellipsis, also has the gap corresponding to a catena. When the gap fails to qualify as a catena, the pseudogapped clause is unacceptable.

(65)



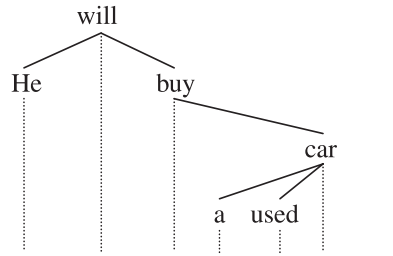
He should leave before Friday, and

a. she should on Friday.

b. *she should on.

[13] Similar data appear to contradict the conclusion, such as *He would study before school if she would after*. Such cases can be explained by acknowledging that certain prepositions can double as pro-form-like adverbs (*before*, *after*, *without*), as in *Do you drink your coffee with cream or without?*

(66)

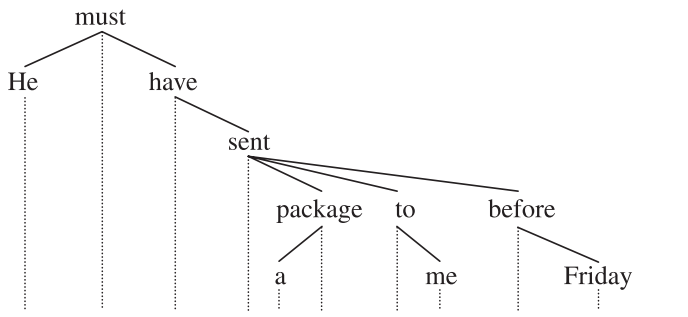


	He	will	buy	a	used	car,	and
a.	she	will		a	used	bike.	
b.	*she	will		a		bike.	

Example (66b) is disallowed on the indicated reading where the bike is necessarily a used bike. The (b)-clauses are disallowed in general because the elided material does not qualify as a catena: the word combination *leave . . . Friday* does not qualify as a catena in (65) and the word combination *buy . . . used* does not qualify as a catena in (66).

The gaps in (65a) and (66a) correspond to just the infinitive in each case. Single words always qualify as strings and as catenae. In this regard, it is not difficult to generate instances of pseudogapping where the gap corresponds to a nonstring word combination.

(67)



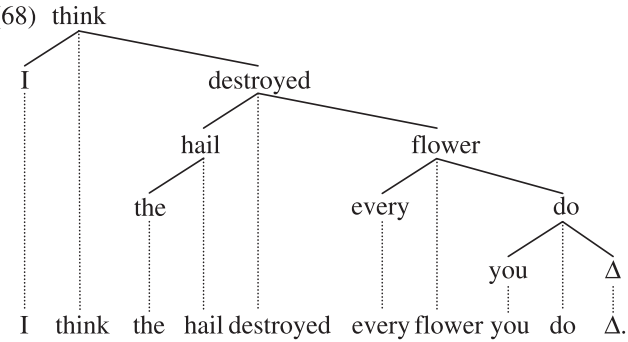
	He	must	have	sent	a	package	to	me	before	Friday, and
a.	she	must	have		a	card.				
b.	she	must	have				to	you.		
c.	she	must	have						on	Friday.
d.	she	must	have		a	card	to	you.		
e.	she	must	have		a	card	to	you	on	Friday.
f.	she	must	have		a	card			on	Friday.
g.	*she	must	have			card.				
h.	*she	must	have				for.			
i.	*she	must	have						on.	
j.	??she	must			a	card.				
k.	*she	must	have					you.		

When the pseudogapped clause is acceptable as in (67a–f), the elided words correspond to a catena. When the pseudogapped words correspond to a noncatena as in (67g–i), the clause is unacceptable. Note also that in each of the acceptable (67a–f), the elided material is *not* a constituent.

Examples (67j) and (67k) are also unacceptable, although the gapped words there do correspond to catenae. Examples (67j) and (67k) thus illustrate an important point about catenae and ellipsis. This point is that the catena requirement on the elided material of pseudogapping and other ellipsis mechanisms is a necessary condition, but not a sufficient one.[14] In other words, it can occur that the elided material corresponds to a catena and the result is nevertheless unacceptable. However, it never occurs that the elided words fail to correspond to a catena and the sentence is nevertheless acceptable.

One final aspect of VP ellipsis is addressed here. The difficulties associated with antecedent-contained deletion (ACD) are immediately overcome given catena-based syntax. The infinite-regress problem that has motivated so many detailed explorations of VP ellipsis (e.g., Bouton 1970; Sag 1976; May 1985; Baltin 1987; Haïk 1987; Larson & May 1990; Fiengo & May 1994; Kennedy 1997; Fox 2002; Harley 2002; Johnson 2001, 2008) does not arise given catenae.

(68) think



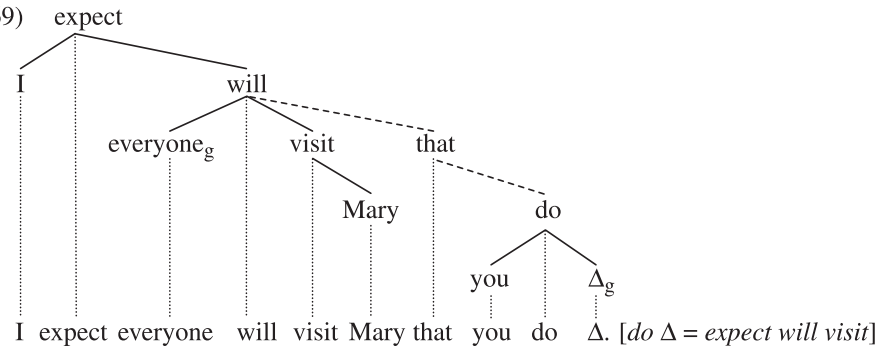
[*do* Δ = *think the hail destroyed*]

The VP ellipsis is marked by *Δ*. Instances of VP ellipsis of this sort are problematic for constituency-based syntax because the apparent antecedent VP to the ellipsis contains the ellipsis itself, which generates an infinite regress.

The preferred solution to the ACD problem is Quantifier Raising (May 1985, Fiengo & May 1994); the quantified expression is raised to a position (at LF) in such a manner that the ellipsis is no longer contained within its antecedent. In contrast, the current system does not appeal to a movement procedure. The antecedent string is straightforwardly a catena; the string *think the hail destroyed* in (68) is a catena. A second example solidifies the point:

[14] Example (67j) is strongly marginal because the status of *must* as an auxiliary verb is questionable. Example (67k) is disallowed because the gap cannot cut into a major constituent. See Hankamer 1973, Neijt 1980, and Osborne 2008:1146ff.

(69)

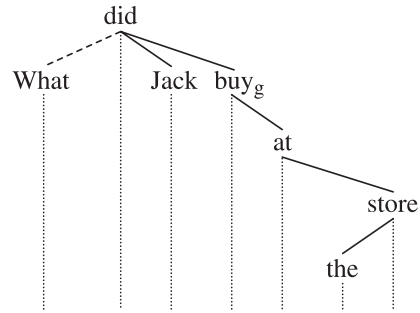


The antecedent to the ellipsis—namely, *expect . . . will visit*—is a catena but clearly not a constituent. There is no need for a movement procedure.[15]

5.5 Sluicing and Comparative Deletion

Sluicing is an ellipsis mechanism that elides (most) everything from a direct or indirect question except for the question word and anything the question word pied-pipes with it (see in this regard Merchant 2001, 2004). The elided material corresponds to a catena:

(70)



A: Jack bought something at the store. What did Jack buy at the store?

B: Really.

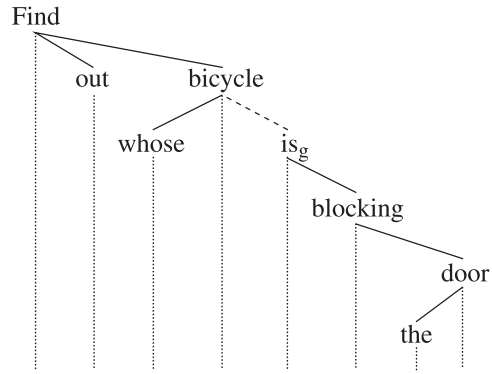
a. What?

b. *What get?

The elided material *did Jack buy at the store* in (70a) corresponds to a catena, whereas the elided material *did Jack . . . at the store* in (70b) does not. The question in (70) is a direct question. Examples of sluicing involving indirect questions also have the elided material as a catena:

[15] Although the ACD problem does not occur in catena-based syntax, other aspects of VP ellipsis pose difficulties. For instance, the position of the governor of the risen relative pronoun within the ellipsis in (69) is mysterious. Furthermore, catena-based syntax contributes no special insights into the phenomenon of *vehicle change* (Fiengo & May 1994).

(71)



A bicycle is blocking the door. Find out whose bicycle is blocking the door.

a. whose.

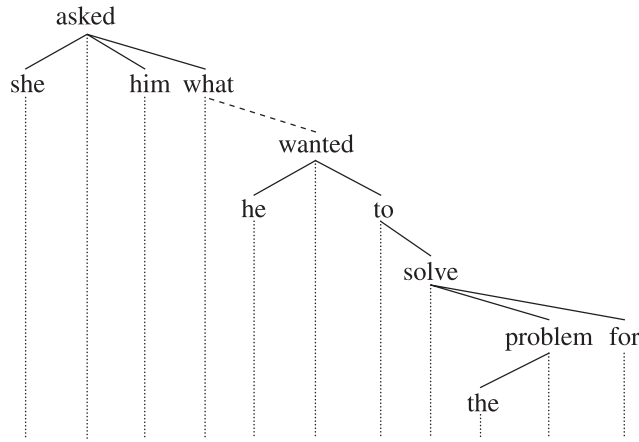
b. whose bicycle.

c. *whose obstructing.

The elided material in both (71a,b) corresponds to a catena. The unacceptable (71c), in contrast, has the elided material corresponding to a noncatena.

Examples (70) and (71) are instances of sluicing where the overt material in the sluiced clause is a catena, for example, *whose bicycle* in (71b). It can occur, however, that the overt material fails to qualify as a catena.

(72)



He wanted to solve the problem, and she asked him what he wanted to solve the problem for.

a. . . . what for.

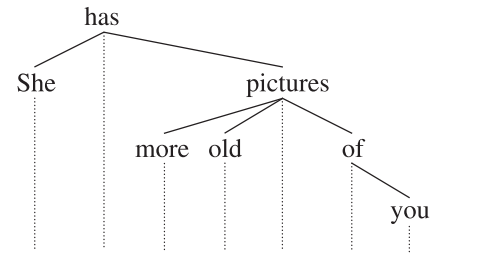
b. . . . *what to do for.

Whereas the remnants *what* and *for* in (72a) do not together form a catena, the elided material *he wanted to solve the problem is* a catena. The sluiced clause in (72b) is bad in part because the sluiced material *he wanted . . . the problem* is not a catena.

Observe also that the elided material *he wanted to solve the problem* is in no way a constituent in (72a), because it excludes the preposition *for*.

The final type of ellipsis mechanism to be considered here is comparative deletion. Comparative deletion usually elides a constituent, but it can elide nonconstituent units as well. When this occurs, the elided material is still a catena.

(73)



	She	has	more	old	pictures	of	you	than
a.	he	has.						
b.	he	has				of	me.	

The elided material in (73a) is the constituent *more old pictures of you*. The elided material *more old pictures* in (73b), however, is not a constituent, yet it is a catena.

The discussion has now demonstrated that the elided material of answer fragments, gapping, stripping, VP ellipsis, pseudogapping, sluicing, and comparative deletion correspond to catenae. Worth emphasizing one last time is that the catena condition on ellipsis is a necessary condition, but not a sufficient one.

6. Conclusion

This paper has presented and defended the claim that the catena (not the constituent) is the fundamental unit of syntax. This claim and the two further claims derived from it are repeated here one last time:

Claim 1: The catena is the fundamental unit of syntax, not the constituent.

Claim 2: All idioms are stored as catenae, but not all idioms are stored as constituents.

Claim 3: The elided material of many ellipsis mechanisms (answer fragments, gapping, stripping, VP ellipsis, pseudogapping, sluicing, comparative deletion) is always a catena, but not always a constituent.

Claims 2 and 3 were defended based on empirical considerations stemming from idioms and the seven ellipsis mechanisms. Given that claims 2 and 3 are derived from claim 1, claim 1 is supported by the empirical considerations that back claims 2 and 3.

The catena is a much more flexible unit of syntax than the constituent. This fact has been demonstrated in terms of the number of catenae and constituents that a given structure contains. Most structures contain many more catenae than they do constituents. It was emphasized in this area that every constituent is a catena, but there are many catenae that are not constituents. It was also emphasized that although

the catena is a flexible unit of syntax, it is also limited insofar as there are many more noncatena word combinations in most structures than there are catena combinations.

By acknowledging the role of catenae, the syntax is linked directly to the semantics. Semantic units that often fail to qualify as constituents are stored as catenae. This fact was illustrated with idioms. The words that form a given idiom often fail to qualify as a constituent at any level of representation or point in a derivation, yet they are always stored as catenae.

The catena is relevant for theories of ellipsis as well. Although the word combinations that many ellipsis mechanisms elide are often nonconstituents, they are always catenae. Thus answer fragments, gapping, stripping, VP ellipsis, pseudogapping, sluicing, and comparative deletion are eliding catenae. When these mechanisms attempt to elide noncatena word combinations, the results are unacceptable. Given this state of affairs, a comprehensive theory of all ellipsis phenomena is now within reach. A fundamental restriction on the named ellipsis mechanisms is easily expressed in terms of catenae.

We believe that the potential of the catena concept is enormous. It can open doors to surface syntactic accounts of many recalcitrant phenomena of the sort touched on in this paper and in other areas. We foresee, for instance, that the domains relevant to binding phenomena can be identified as catenae and that meaning bearing morph combinations inside and between words also form catenae. Further research will determine the extent to which the catena concept is useful to these and other areas.

Finally, the catena has been explored here using our dependency-based grammar. The catena concept is, however, not limited to just dependency grammars but rather it can be defined over constituency-based structures as well. In this regard we invite readers to adopt the concept into their preferred frameworks and to test its applicability.

References

Ágel, V., , L. Eichinger, H.-W. Eroms, P. Hellwig, H. J. Heringer & H. Lobin, eds. 2003–2006. *Dependency and valency: An international handbook of contemporary research*. Berlin: Walter de Gruyter.

Baltin, M. 1987. Do antecedent-contained deletions exist? *Linguistic Inquiry* 18:579–595.

Bouton, L. 1970. Antecedent-contained pro-forms. In *Proceedings of the sixth regional meeting of the Chicago Linguistic Society*, ed. M. Campbell, 154–167. Chicago: University of Chicago Press.

Brody, M. 2000. Mirror Theory: Syntactic representation in Perfect Syntax. *Linguistic Inquiry* 31:29–56.

Bröker, N. 2000. Unordered and non-projective dependency grammars. In *Les grammaires de dépendance* [Dependency grammars] (Traitement automatique des langues 41), ed. S. Kahane, 79–111. Paris: Hermes.

Bröker, N. 2003. Formal foundations of dependency grammar. In *Dependency and valency: An international handbook of contemporary research*, ed. V. Ágel, L. Eichinger, H.-W. Eroms, P. Hellwig, H. J. Heringer & H. Lobin, vol. 1, 294–310. Berlin: Walter de Gruyter.

Chomsky, N. 1957. *Syntactic structures*. The Hague: Mouton and Co.

Duchier, D. & R. Debusmann. 2001. Topology dependency trees: A constraint-based account of linear precedence. In *39th annual meeting of the Association for Computational*

Linguistics: Proceedings from the conference, 180–187. Stroudsburg, PA: Association for Computational Linguistics.
Embick, D. & R. Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32:555–595.
Engel, U. 1994. *Syntax der deutschen Gegenwartssprache* [Syntax of modern German]. 3rd ed. Berlin: Erich Schmidt.
Eroms, H.-W. 1985. Ein reine Dependenzgrammatik für das Deutsche [A pure dependency grammar for German]. *Deutsche Sprache* 13:306–326.
Eroms, H.-W. 2000. *Syntax der deutschen Sprache* [Syntax of the German language]. Berlin: Walter de Gruyter.
Eroms, H.-W. & H. J. Heringer. 2003. Dependenz und lineare Ordnung [Dependency and linear order]. In *Dependency and valency: An international handbook of contemporary research*, ed. V. Ágel, L. Eichinger, H.-W. Eroms, P. Hellwig, H. J. Heringer & H. Lobin, vol. 1, 247–262. Berlin: Walter de Gruyter.
Fiengo, R. & R. May. 1994. *Indices and identity*. Cambridge, MA: MIT Press.
Fillmore, C., P. Kay & M. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64:501–538.
Fox, D. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33:63–96.
Goldberg, A. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University Press of Chicago.
Goldberg, A. 2006. *Constructions at work*. Oxford: Oxford University Press.
Groß, T. 1999. *Theoretical foundations of dependency syntax*. Munich: Iudicium.
Groß, T. 2010. Chains in syntax and morphology. In *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation at Tohoku University*, ed. O. Ryo, K. Ishikawa, H. Uemoto, K. Yoshimoto & Y. Harada, 143–152. Tokyo: Waseda University.
Groß, T. & T. Osborne. 2009. Toward a practical DG theory of discontinuities. *Sky Journal of Linguistics* 22:43–90.
Haïk, I. 1987. Bound VPs that need to be. *Linguistics and Philosophy* 10:503–530.
Halle, M. & A. Marantz. 1993. Distributed Morphology and the pieces of inflection. In *The view from Building 20*, ed. K. Hale & S. Keyser, 111–176. Cambridge, MA: MIT Press.
Hankamer, J. 1973. Unacceptable ambiguity. *Linguistic Inquiry* 4:17–68.
Harley, H. 2002. ACO, ACD, and QR of DPs. *Linguistic Inquiry* 33:659–664.
Hays, D. 1964. Dependency theory: A formalism and some observations. *Language* 40:511–525.
Hellwig, P. 2003. Dependency Unification Grammar. In *Dependency and valency: An international handbook of contemporary research*, ed. V. Ágel, L. Eichinger, H.-W. Eroms, P. Hellwig, H. J. Heringer & H. Lobin, vol. 1, 593–635. Berlin: Walter de Gruyter.
Heringer, H. J. 1996. *Deutsche Syntax Dependentiell* [German syntax as dependencies]. Tübingen, Germany: Staufenberg.
Horn, G. 2003. Idioms, metaphors, and syntactic mobility. *Journal of Linguistics* 39:245–273.
Hudson, R. 1984. *Word Grammar*. New York: Basil Blackwell.
Hyvärinen, I. 2003. Der verbale Valenzträger. In *Dependency and valency: An international handbook of contemporary research*, ed. V. Ágel, L. Eichinger, H.-W. Eroms, P. Hellwig, H. J. Heringer & H. Lobin, vol. 1, 738–763. Berlin: Walter de Gruyter.
Jackendoff, R. 1971. Gapping and related rules. *Linguistic Inquiry* 2:21–35.
Jayaseelan, K. A. 1990. Incomplete VP deletion and gapping. *Linguistic Analysis* 20:64–81.
Johnson, K. 2001. What VP ellipsis can do and what it can't, but not why. In *The handbook of contemporary syntactic theory*, ed. M. Baltin & C. Collins, 439–474. Oxford: Blackwell.
Johnson, K. 2008. The view of QR from ellipsis. In *Topics in ellipsis*, ed. K. Johnson, 69–94. Cambridge: Cambridge University Press.
Jung, W.-Y. 1995. *Syntaktische Relationen im Rahmen der Dependenzgrammatik*. Hamburg: Buske.
Kahane, S., ed. 2000. *Les grammaires de dépendance* [Dependency grammars] (Traitement automatique des langues 41). Paris: Hermes.

Kay, P. & C. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction. *Language* 75:1–33.
Kennedy, C. 1997. Antecedent-contained deletion and the syntax of quantification. *Linguistic Inquiry* 28:662–688.
Koopman, H. & D. Sportiche. 1991. The position of subjects. *Lingua* 85:211–258.
Kuno, S. 1981. The syntax of comparative clauses. In *Papers from the 17th regional meeting: Chicago Linguistic Society*, ed. R. Hendrick, C. Masek & F. Miller, 136–155. Chicago: Chicago Linguistic Society.
Kunze, J. 1975. *Abhängigkeitsgrammatik* [Dependency Grammar] (Studia Grammatica 12). Berlin: Akademie Verlag.
Lakoff, G. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
Larson, R. 1988. On the double object construction. *Linguistic Inquiry* 19:335–391.
Larson, R. & R. May. 1990. Antecedent containment or vacuous movement: Reply to Baltin. *Linguistic Inquiry* 21:103–122.
Lasnik, H. 1999. Pseudogapping puzzles. In *Fragments: Studies in ellipsis and gapping*, ed. S. Lappin & E. Benmamoun, 141–174. New York: Oxford University Press.
Lobin, H. 1993. *Koordinationssyantax als prozedurales Phänomen* [The syntax of coordination as a procedural phenomenon] (Studien zur deutschen Grammatik 46). Tübingen, Germany: Narr.
Matthews, P. 1981. *Syntax*. Cambridge: Cambridge University Press.
May, R. 1985. *Logical Form: Its structure and derivation*. Cambridge, MA: MIT Press.
Mel'čuk, I. 1988. *Dependency syntax: Theory and practice*. Albany, NY: State University of New York Press.
Merchant, J. 2001. *The syntax of silence: Sluicing, islands, and identity in ellipsis*. New York: Oxford University Press.
Merchant, J. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27:661–738.
Neijt, A. 1980. *Gapping: A contribution to sentence grammar*. Dordrecht, the Netherlands: Foris.
Nunberg, G., I. Sag & T. Wasow. 1994. Idioms. *Language* 70:491–538.
O'Grady, W. 1998. The syntax of idioms. *Natural Language & Linguistic Theory* 16:79–312.
Osborne, T. 2005. Beyond the constituent: A dependency grammar analysis of chains. *Folia Linguistica* 39:251–297.
Osborne, T. 2006. Shared material and grammar: A dependency grammar theory of non-gapping coordination. *Zeitschrift für Sprachwissenschaft* 25:39–93.
Osborne, T. 2007. The weight of predicates: A dependency grammar analysis of predicate weight in German. *Journal of Germanic Linguistics* 19:23–72.
Osborne, T. 2008. Major constituents: And two dependency grammar constraints on sharing in coordination. *Linguistics* 46:1109–1165.
Pickering, M. & G. Barry. 1993. Dependency Categorial Grammar and coordination. *Linguistics* 31:855–902.
Robinson, J. 1970. Dependency structures and transformational rules. *Language* 46:259–285.
Sag, I. 1976. Deletion and Logical Form. Ph.D. dissertation, MIT, Cambridge, MA.
Schubert, K. 1988. *Metataxis: Contrastive dependency syntax for machine translation*. Dordrecht, the Netherlands: Foris.
Siewierska, A. 1988. *Word order rules*. London: Croom Helm.
Sgall, P., E. Hajičová & J. Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Dordrecht, the Netherlands: D. Reidel.
Sportiche, D. 2005. Division of labor between merge and move: Strict locality of selection and apparent reconstruction paradoxes. Ms., University of California, Los Angeles. http://ling.auf.net/lingBuzz/000163.
Starosta, S. 1988. *The case for Lexicase: An outline of Lexicase grammatical theory*. New York: Pinter Publishers.

Tarvainen, K. 2000. *Einführung in die Dependenzgrammatik* [Introduction to Dependency Grammar]. Tübingen, Germany: Niemeyer.
Tesnière, L. 1959. *Éléments de syntaxe structurale* [Elements of structural syntax]. Paris: Klincksieck.
Tesnière, L. 1969. *Éléments de syntaxe structurale* [Elements of structural syntax]. 2nd ed. Paris: Klincksieck.

Timothy Osborne
3545 La Fontana Dr.
Boise, ID 83702
USA

tjo3ya@yahoo.com

Michael Putnam
Pennsylvania State University
Department of Germanic Languages and Literatures
427 Burrowes Building
University Park, PA 16802
USA

mtp12@psu.edu

Thomas Groß
Aichi University
Department of Language Communication
Faculty of International Communication
441-8522 Machihata-cho 1-1, Toyohashi-shi, Aichi-ken
Japan

tmgross@vega.aichi-u.ac.jp