

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

COGNITION

Cognition 105 (2007) 300–333

[www.elsevier.com/locate/COGNIT](http://www.elsevier.com/locate/COGNIT)

# Minimization of dependency length in written English <sup>☆</sup>

David Temperley <sup>\*</sup>*Eastman School of Music, University of Rochester, Rochester, NY, USA*

Received 10 November 2005; revised 12 September 2006; accepted 18 September 2006

---

## Abstract

Gibson's Dependency Locality Theory (DLT) [Gibson, E. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1–76; Gibson, E. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain* (pp. 95–126). Cambridge, MA: MIT Press.] proposes that the processing complexity of a sentence is related to the length of its syntactic dependencies: longer dependencies are more difficult to process. The DLT is supported by a variety of phenomena in language comprehension. This raises the question: Does language *production* reflect a preference for shorter dependencies as well? I examine this question in a corpus study of written English, using the Wall Street Journal portion of the Penn Treebank. The DLT makes a number of predictions regarding the length of constituents in different contexts; these predictions were tested in a series of statistical tests. A number of findings support the theory: the greater length of subject noun phrases in inverted versus uninverted quotation constructions, the greater length of direct-object versus subject NPs, the greater length of postmodifying versus premodifying adverbial clauses, the greater length of relative-clause subjects within direct-object NPs versus subject NPs, the tendency towards “short-long” ordering of postmodifying adjuncts and coordinated conjuncts, and the shorter length of subject NPs (but not direct-object NPs) in clauses with premodifying adjuncts versus those without.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Linguistic complexity; Syntactic dependency; Sentence processing; Corpus studies

---

<sup>☆</sup> This manuscript was accepted under the editorship of Jacques Mehler.

<sup>\*</sup> Tel.: +1 585 274 1557; fax: +1 585 274 1088.

E-mail address: [dtemperley@esm.rochester.edu](mailto:dtemperley@esm.rochester.edu)

## 1. Dependency length and complexity

An important element of modern linguistic theory is the concept of *dependencies*. A dependency is an asymmetrical syntactic relation between a pair of words in a sentence (known as the *head* and the *dependent*); the head of each dependency is then the dependent of another word (unless it is the head of the entire sentence), forming a recursive structure which connects the entire sentence. In many linguistic theories, dependencies play an important role (Bresnan, 1982; Dik, 1989; Hudson, 1990; Mel'cuk, 1987; Oehrle, Bach, & Wheeler, 1988; Pollard & Sag, 1987; Radford, 1997). For the most part, there is general agreement as to the nature of dependency structures in language. In the case of prepositional phrases in English, for example, it is usually assumed that the preposition is the head of the phrase (with the prepositional object as its dependent) and is then a dependent of the word (generally a preceding verb or noun) that the phrase conventionally modifies. In general, the head of each major constituent type (NP, VP, AP, PP) is the word after which the phrase is named, and the head of a clause is its finite verb.<sup>1</sup>

The concept of dependencies has also proven to have great explanatory value in psycholinguistics. Recent work, especially that of Gibson (1998, 2000), has shown that the complexity of processing a sentence is directly related to the length of the dependencies within it. According to Gibson's Dependency Locality Theory (DLT), the syntactic complexity of a sentence can be predicted by two factors: "storage cost", the cost of maintaining in memory the syntactic predictions or requirements of previous words; and "integration cost", the cost of syntactically connecting a word to previous words with which it has dependent relations.<sup>2</sup> The integration cost for a word increases with the distance to the previous words with which it is connected, on the reasoning that the activation of words decays as they recede in time, making integration more difficult. Gibson shows that the DLT predicts a number of phenomena in comprehension, such as the greater complexity of object-extracted versus subject-extracted relative clauses (King & Just, 1991). In both subject relatives (1a) and object relatives (1b), the verb of the relative clause (*attacked*) is dependent on the preceding relative pronoun (*who*); in subject relatives, these two words are normally adjacent, but in object relatives they are separated by the relative clause subject (*the senator*), yielding a higher integration cost.

<sup>1</sup> One controversial case is noun phrases; most theories assume that NPs are headed by their main nouns (Bresnan, 1982; Gibson, 1998; Mel'cuk, 1987; Pollard & Sag, 1987), but recent theory in the GB/minimalist vein assumes the determiner as the head (Abney, 1987; Radford, 1997). We will assume the main noun as head of the NP, though none of the results presented here hinge on this issue. Coordinate structures are also a problematic case, as I discuss below.

<sup>2</sup> The DLT is presented in Gibson (2000) and is a modified version of a theory presented earlier, the Syntactic Prediction Locality Theory (SPLT) (Gibson, 1998). The two theories are similar, the main difference being that in the SPLT, storage cost (also called "memory cost") increases with the distance between words, while in the DLT it does not. (Gibson (2000) argues that if integration cost is distance-based, then storage cost need not be; the fact that the activation of words decreases with distance—thus increasing integration cost—indirectly captures the added storage cost due to intervening words.) Here, we focus on integration cost, which increases with word distance in both versions of the theory.

(1)

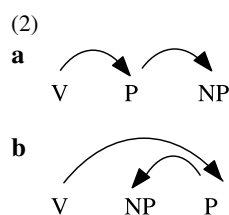
a. The reporter who attacked the senator admitted the error.

b. The reporter who the senator attacked admitted the error.

The theory also predicts phenomena in ambiguity resolution: in prepositional-phrase attachment decisions (Gibson & Pearlmutter, 1994; Thornton, MacDonald, & Arnold, 2000) and main-verb/reduced-relative ambiguities (Gibson, 2000), the preferred interpretation reflects a preference for structures with shorter dependencies.

The DLT follows in a long tradition of theories that posit distance between related words as an important factor in sentence processing and complexity (Behagel, 1932; Frazier, 1985; Hawkins, 1994). To advocate this view is not, of course, to deny that other factors may affect sentence complexity; but Gibson argues that the DLT accounts for a variety of important phenomena not explained by other theories. Gibson's focus has mainly been on phenomena of comprehension, showing that structures with shorter dependencies are preferred and easier to process. We might also ask: does language *production* reflect a preference for structures with shorter dependencies? This possibility is discussed briefly by Gibson (1998, p. 52), but has not really been explored. Other authors, notably Hawkins (1994, 2004), have argued that considerations of parsing complexity affect production, and have found evidence for this in corpus data; it seems reasonable to ask whether the DLT is supported by production evidence as well.<sup>3</sup> In this study, I examine this question in a corpus study of written English, to see if language users tend to favor structures with shorter dependencies.

My main concern here is with phenomena of syntactic *choice*, not those that are dictated by actual grammatical rules. To illustrate this point, consider the case of adpositional phrases modifying verbs. (*Adposition* is the general term for prepositions, which precede their objects, as in English, and postpositions, which follow them, as in Japanese.) In cases where the adpositional phrase follows the verb—as it normally does in English—it can be seen that dependency lengths will be shorter if the adposition (P) precedes its object NP, as in (2a) below, rather than following it, as in (2b).



However, this is not a phenomenon of syntactic choice (at least not in English). English speakers do not have the option of placing the adposition after its object; to do so would be grammatically incorrect. The fact that the English language is prepositional and not postpositional may reflect pressure towards dependency-length

<sup>3</sup> Hawkins's EIC theory is closely related to the DLT, in that it relates complexity to the proximity of related words; however, the two theories differ in certain key respects. Hawkins' theory will be discussed at some length in Section 4.

minimization (we return to this point below); but this is an issue of linguistic evolution, not syntactic choice. My focus here, rather, will be on phenomena that could be seen as reflecting spontaneous choices among users of a language, not simply adherence to hard-and-fast grammatical rules.

To determine if writers favor structures with shorter dependencies, we need to know: What kind of syntactic structures yield shorter dependency lengths? In what follows, I present four principles or “Dependency Length Minimization Rules” (DLMRs) stating the factors that minimize dependency length. These rules lead to predictions about the relative length of constituents in English: for example, subject vs. object noun-phrases. For each prediction, I present a statistical test comparing the lengths of constituent types in different contexts to see if the prediction is confirmed. (The length of a constituent is defined simply as the number of words it contains.<sup>4</sup>) The tests all use the Wall Street Journal portion of the Penn Treebank (Marcus et al., 1994), a corpus of over 1 million words of English text from the 1989 Wall Street Journal, annotated with syntactic information. (Hereafter I will simply call this the “WSJ corpus”.)

Before continuing, we should note that the theoretical proposal under consideration here—that the syntactic complexity of a sentence (or phrase) is related to its total dependency length—may be slightly different from the DLT, in two ways. The first concerns the measurement of dependency length. We assume here that the length of a dependency corresponds to the number of words spanned, so that (for example) a dependency connecting adjacent words has a length of 1. By contrast, Gibson (1998, 2000) defines the distance between two words in terms of the number of intervening “new discourse referents.” However, this aspect of the DLT seems provisional: Gibson states, “[i]t is also likely that processing every intervening word, whether introducing a new discourse structure or not, causes some integration cost increment” (1998, p. 13). A second difference concerns the concept of complexity. Gibson suggests that the “intuitive complexity” of a sentence depends on its *maximal* integration cost (Gibson, 1998, pp. 16–17; Gibson, 2000, p. 105). An integration cost is incurred when a word connects to a previous word or words, and is equal (or proportional) to the total length of all dependencies being formed at that point.<sup>5</sup> For Gibson, then, intuitive complexity depends on the maximum value of this cost at any point in a sentence. However, one could also define complexity in terms of processing time or computational effort—and in some cases Gibson does exactly this, for example, using integration costs to predict word-by-word reading times in relative clauses (Gibson, 2000). In this case, it seems clear that the total complexity of a sentence (or part of a sentence) is not given by the maximal integration cost, but rather should reflect the total of all integration costs. Since each dependency ultimately contributes to the integration cost of a word (the word on its right end),

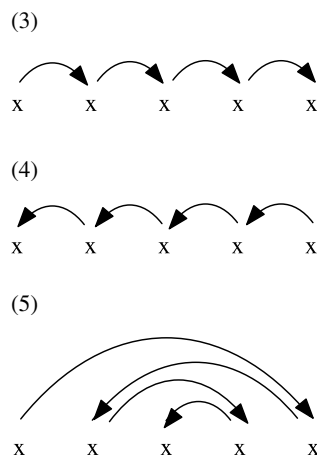
<sup>4</sup> In measuring the length of constituents, I excluded non-lexical punctuation symbols like periods and commas (but included lexical ones like \$ and %). Beyond this, I followed the tokenization of words indicated in the Penn Treebank. Thus contractions like *don't* were treated as two words, as were possessives like *John's*.

<sup>5</sup> In Gibson (2000), the integration cost of a dependency is defined to be linearly related its length—the number of discourse referents spanned (though Gibson acknowledges that this may be oversimplified). I adopt this assumption here as well (while measuring length in words rather than discourse referents).

the total integration cost of a sentence is simply equal to the sum total of all its head-dependent distances. In the current study, then, complexity is measured as the total length of all dependencies in a phrase or sentence.<sup>6</sup> Notwithstanding these small differences, the current proposal is clearly closely related to the Dependency Locality Theory, and it seems fair to regard the tests presented below as tests of that theory.

## 2. Preference for short left-branching constituents

If we consider a simple structure in which each word has exactly one dependent, we can see that dependency length is minimized if each word's dependent is in the opposite direction from its head: to put it another way, the words should either be all “right-branching”, as in (3) below, or all “left-branching”, as in (4). (A dependent word is right-branching if it is to the right of its head, left-branching if it is to the left. Similarly, a constituent is right-branching if it is to the right of the *external* head to which it connects, left-branching if it is to the left.) A mixture of branching, as in (5), can cause longer dependency lengths. (In all diagrams, dependencies are indicated by arrows pointing from the head to the dependent.)



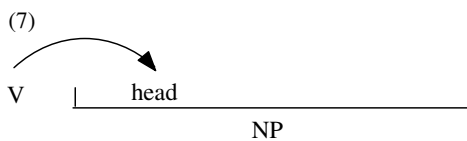
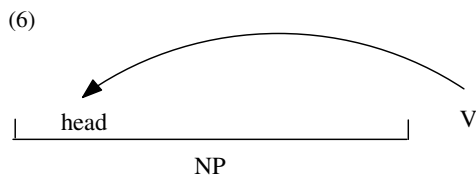
<sup>6</sup> Another issue requiring discussion is “empty categories”. The DLT (like many current syntactic theories) assumes syntactic structures in which certain constituents contain no overt lexical items—for example, the subject NP in a subject-extracted relative clause. Empty categories pose a difficult problem, since there is little agreement as to their location or even their existence (Pickering & Barry, 1991). However, Gibson (1998, p. 21), points out that the inclusion of empty categories may not greatly affect the predictions of the DLT. Consider the case of NP gaps in subject relative clauses—a sentence such as *The reporter who attacked the senator admitted the error*. One could either say an empty relative-clause subject (after the relative pronoun *who*) attaches to the relative pronoun, or one could attach the verb *attacked* to the relative pronoun directly; the difference in dependency lengths is very small. In other cases—e.g. with prepositional relatives—Gibson notes that complexity judgments may argue against the existence of empty categories (1998, pp. 44–5). In this study, we assume syntactic structures with no empty categories. While some empty categories are represented in the Penn Treebank, these are ignored in the tests reported below; null lexical items are not counted in measuring the length of constituents, and empty constituents are not counted at all (the number of constituents of length zero is always assumed to be zero).

This leads to our first rule:

**DLMR 1.** Dependency structures should be either consistently right-branching or left-branching.

It is generally agreed that languages tend to obey this rule: they tend to be either consistently “head-first” or “head-last” (Chomsky, 1988; Hawkins, 1983; Vennemann, 1974). For example, languages (like English) in which verbs precede their objects (and other complements) tend to be “prepositional”, with adpositions also preceding their objects. Some authors have suggested that the general tendency for languages to be consistently head-first or head-last may be due to the pressure to keep dependents close to their heads (Frazier, 1985; Rijkhoff, 1990). This is primarily a matter of syntactic *rules* (grammar), not syntactic choice, and thus does not really concern us here.

While languages tend to be either primarily left-branching or right-branching, many languages are not entirely consistent in this regard. English is mainly right-branching, but has left-branching structures in certain situations. For example, NPs are right-branching in the case of direct objects (connecting to the preceding main verb), but left-branching in the case of subjects (connecting to the following finite verb). If a language is mostly right-branching, the dependents within each constituent C will tend to be to the right of the head of C; thus the head is likely to be at or near the left end of C. With regard to English noun-phrases, for example, the head of an NP is usually near its left end when the phrase is long (evidence for this will be given below). What is of interest is the distance between the head of a constituent C and the external head of which it is a dependent—what we will call the “parent head” of C. If an NP is left-branching (e.g. a subject NP phrase) and long, there may be a large distance between the NP head and the parent head (V); see (6) below. If the NP is right-branching (e.g. a direct-object phrase), it may be long without greatly lengthening the dependency from the head to the parent head (7).



One could avoid undesirable situations like (6) by making left-branching constituents short. We express this in a second rule:

DLMR 2. In a primarily right-branching language, the left-branching constituents should be short.

A fairly straightforward test of DLMR 2 concerns subject-verb inversion in quotation constructions. In general, the subject precedes the finite verb in declarative sentences, and follows it in most types of questions. In quotation constructions, however—where a quotation is followed by a quoting verb—there is a choice: one may either invert the verb and subject (8a and 8c) or not (8b and 8d).

(8)

- a. “I agree”, [said [Jane]].
- b. “I agree”, [[Jane] said].
- c. “I agree”, [said [Jane Smith, president of Smith, Brown, & Jones, a consulting firm]].
- d. “I agree”, [[Jane Smith, president of Smith, Brown, & Jones, a consulting firm,] said].

In (8a) and (8c), the subject of the quoting verb *said* is right-branching in relation to the verb; in (8b) and (8d), it is left-branching. If the subject phrase is long, as in (8c) and (8d), a left-branching construction such as (8d) creates a long dependency between the head of the subject NP (*Smith*) and the following verb; this can be avoided by using an inverted construction like (8c). The prediction of DLMR 2, then, is that long subject phrases will more often be used in inverted structures like (8c) (“V–S” order) than in non-inverted structures like (8d) (“S–V” order); if the subject NP is short, the pressure in favor of V–S order is absent. The average length of subject NPs should therefore be greater in V–S constructions.

Test 1: Subject NPs in quotation sentences, in S–V order and V–S order

Result: Average length of subject NP: in S–V order, 2.16 words; in V–S order, 9.47 words. ( $F(1,4071) = 53520.6, p < .0005$ )

(The appendix gives details as to how inverted and non-inverted subject NPs and other constituent types discussed below were defined in terms of Penn Treebank notation.) The prediction is confirmed: Subject NPs are much longer in V–S constructions than in S–V constructions.

The argument just presented hinges on the assumption that the heads of NPs tend to be near their beginnings, particularly if the NP is long. If the head of the subject NP in (8c) and (8d) were near the end of the NP, then it would actually be closer to the verb in (8d) than in (8c). (If the subject NP is short—e.g. 2 words

long—it will not have a long dependency to its parent head in either S–V or V–S order, regardless of the position of the head within the NP.) When an NP is more than a few words long, this is usually because it contains some kind of modifying phrase—such as a prepositional phrase, relative clause, or appositive—and these almost invariably follow the head noun. Thus the assumption that the head of a long NP is usually near the beginning seems plausible. It seemed prudent to test this assumption, however, and this was done in a corpus analysis, using an algorithm created by Collins (1999) for identifying dependencies in Penn Treebank data. All NPs in the WSJ corpus were identified along with their heads and the position of the head in the NP (with 0.0 being the first word and 1.0 being the last word). (In 1-word constituents, the position of the word was defined as 0.5.) NPs were grouped by length and the average head position was calculated for each length. The results are shown in Fig. 1. It can be seen that, indeed, for long NPs (though not for short NPs), the head tends to be near the beginning. Consider 11-word phrases; in this case, the words are at positions 0.0, 0.1, 0.2 and so on up to 1.0. The average head position in this case is at 0.173, between words 2 and 3. Thus the assumption of the previous test is confirmed.

A further prediction of DLMR 2 concerns subject and direct-object NPs. As noted earlier, subject NPs in English are generally left-branching while direct-object NPs are right-branching. The prediction, then, is that long subject NPs will be avoided. (If we imagine the NP in (6) as a subject NP and the one in (7) as an object NP, a long NP creates a long dependency in (6) but not in (7).) If this is the case, the mean length of subject NPs should be shorter than that of direct-object NPs. (In this test we consider only subject NPs in uninverted “S–V” order.)

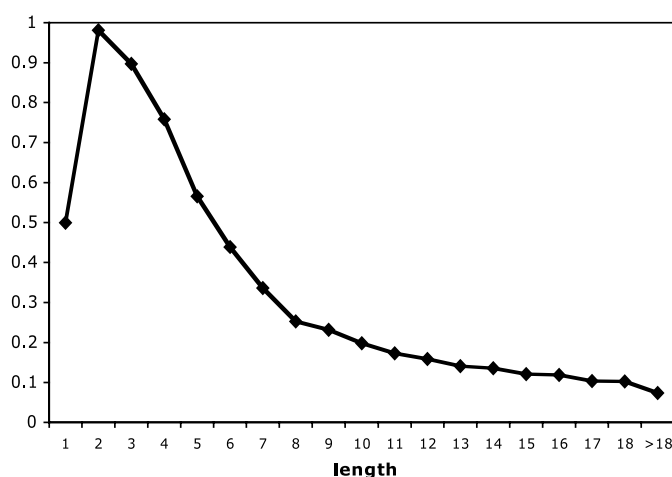


Fig. 1. Average position of the head in NPs of different lengths, where 0.0 is the first word and 1.0 is the last word.



Test 2: Subject NPs (in S–V constructions) versus direct-object NPs

Result: Average length of subject NPs = 3.13 words; average length of direct-object NPs = 5.80 words. ( $F(1, 124534) = 10385.7, p < .0005$ )

The prediction is confirmed: direct-object NPs are significantly longer than subject NPs.

In comparing subject and direct-object NPs, we are extending the idea of “syntactic choice” in a way that requires some discussion. In the case of subject–verb inversion in quotation constructions, there is a choice between two orderings of the same words that seem equivalent semantically (and perhaps even pragmatically), and it seems plausible that this choice is at least partially determined by processing considerations. In the case of subject versus object NPs, there is no such choice in the ordering of constituents within the sentence. (Obviously, simply swapping the subject and object of a sentence is not usually an option as it would change the meaning of the sentence.) Still, languages users do have choices as to how they construct a sentence or a longer passage of discourse; there are usually many ways of saying the same thing. In the case of subjects and objects, for example, one can often express essentially the same thought with a particular noun phrase in either subject or object position. Consider this sentence from the WSJ corpus:

- (9) For years, a strict regimen governed [NP the staff meetings at Nissan Motor Co.’s technical center in Tokyo’s Western suburbs].

The sentence ends with a long NP in direct-object position. One might well have phrased the sentence with the long NP in subject position:

- (10) For years, [NP the staff meetings at Nissan Motor Co.’s technical center in Tokyo’s western suburbs] followed a strict regimen.

Perhaps the author avoided this option, in part, because it would have resulted in a long subject NP. Another case in point concerns modifying phrases such as non-restrictive relative clauses and appositives—phrases adding elaborative information to a noun phrase (but not information that is essential to the identification of the referent, as in a restrictive relative clause). In cases where a discourse entity is mentioned several times, the writer often has a choice as to where these modifying phrases are placed. Consider this passage from the WSJ corpus (the passage contains the first three sentences of a news article):

- (11) CONGRESSIONAL LEADERS BACKED Bush’s criticism of Nicaragua’s Ortega. While lawmakers haven’t raised the possibility of renewing military aid to the Contras following Ortega’s weekend threat to end a truce, Senate Majority Leader Mitchell said on NBC-TV that Ortega

had made “a very unwise move.” Minority Leader Dole plans to offer a resolution tomorrow denouncing the Nicaraguan president, whose remarks came during a celebration in Costa Rica marking regional moves to democracy.

The discourse contains four references to Nicaraguan president Daniel Ortega—three times as *Ortega*, the fourth as *the Nicaraguan president*. The fourth reference is elaborated by a non-restrictive possessive relative clause: *whose remarks came during a celebration in Costa Rica marking regional moves to democracy*. This relative clause could perfectly well have been placed in the second sentence, after the third reference to Ortega:

- (12) ...Senate Majority Leader Mitchell said on NBC-TV that Ortega, whose remarks came during a celebration in Costa Rica marking regional moves to democracy, had made “a very unwise move.”

This alternative way of structuring the discourse is identical in meaning to the original; even in pragmatic or informational terms, it is difficult to see how it is less felicitous than, or even different from, the original. But (12) creates a very long dependency between the embedded clause subject *Ortega* and the verb *had*; perhaps the author avoided it for this reason. This could well be regarded as a kind of syntactic choice, as there are various ways of constructing the discourse (various possible locations for the relative clause) which are different in syntax but essentially the same semantically and pragmatically. Perhaps the greater length of object NPs relative to subject NPs is due to the fact that they more often are given elaborative information of this sort.

One might offer other explanations for the length difference between subject and direct-object phrases—in particular, the “given-new” distinction. An important concept in discourse analysis is the distinction between entities that are given (previously mentioned in the discourse) and those that are new (Clark & Haviland, 1977; Gundel, Hedberg, & Zacharski, 1993). New discourse entities presumably tend to be longer, and it has also been suggested that they more often occur in object position (Branigan, McLean, & Reeve, 2003); thus the greater length of object phrases might be due to the fact that they are more often new. One way to control for this would be to compare the length of subject and object NPs, looking only at NPs representing new discourse entities. How can this be done? As a starting point, we could compare NPs beginning with the indefinite article *a/an*. However, while many such NPs represent new discourse items (so-called “specific” indefinite NPs), such as (13a) below, others do not. Non-specific uses of indefinite NPs include generic NPs (13b), modal statements such as some conditional, intentional, or predictive state-

ments (13c), and statements in which the indefinite NP is within the scope of another quantifier (13d).<sup>7</sup>

(13)

- a. *A company spokesman* declined to elaborate on the departure.
- b. *A soft landing* is an economic slowdown that eases inflations without leading to a recession.
- c. I'm afraid *a jury* would not have convicted her.
- d. . . .many bad movies have *a bright spot* . . .

To avoid possible confounds—for example, the possibility that generic NPs might occur more often in subject position—we must ensure that only specific indefinite NPs are considered. Since these are not explicitly identified in the WSJ corpus, they had to be identified by hand. All subject and direct-object NPs in the corpus beginning with *a/an* were extracted, and were examined one by one (starting from the beginning of the corpus) and classified as specific and non-specific, until 100 specific subject NPs and 100 specific direct-object NPs were found; these two sets were then compared with regard to their average word length.

<sup>7</sup> While a variety of semantic and pragmatic treatments of indefinite NPs have been proposed, most agree on a distinction roughly corresponding to that proposed here between “specific” indefinite NPs and others (Abbott, 2004; Diesing, 1992; Heim, 1983; Partee, 1972). In traditional semantics, a specific NP is one with an existential quantifier taking “broad scope”: *A spokesman said X* is analyzed as *There exists a spokesman such that the spokesman said X*. Other indefinite NPs include those where there is no existential quantifier (*A dog has four legs* is analyzed as *For every dog, if X is a dog X has four legs*), or where the existential quantifier takes “narrow scope” within another quantifier or modal operator (*Every movie has a bright spot* is analyzed as *For every bad movie X, there exists a bright spot Y such that X has Y*). A different approach—but leading to similar results—is that of Heim (1983), who argues that specific indefinite NPs are used to introduce a new entity or “file card” into the discourse. Sentences such as (13b), (13c), and (13d) do not introduce new discourse entities, but rather imply general statements about a class of entities. I used the following diagnostic tests to determine whether the phrase *an X* in the sentence [S . . . *an X* etc.] was a specific indefinite NP:

1. If one can rephrase the sentence as *There exists an X such that* [S . . . *the X* etc.], it is a specific NP.
2. If, following *an X*, subsequent mentions of this discourse entity seem to require *the X* rather than *a/an X*, the initial expression is a specific NP.
3. If *an X* can be replaced by (a) *any X*, or (b) *an X, as yet undetermined*, it is a non-specific NP.

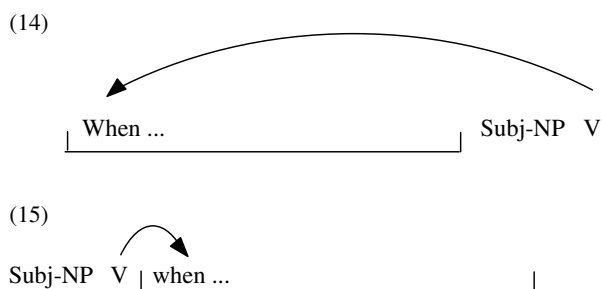
While not perfect, these three tests lead to clear and convergent conclusions in a vast majority of cases. There were a few borderline cases; most concerned statements of intention in which it was not clear whether the NP was specific or individuated enough to be considered a new discourse entity. An example from the WSJ corpus is the sentence *Saudi Arabia . . . has vowed to enact a copyright law compatible with international standards*. If this only represents a vague intention concerning the nature of the copyright law to be enacted, it might be considered a non-specific statement in which the indefinite NP takes narrow scope: *Saudi Arabia intends that there will exist a copyright law compatible with international standards*. But if the copyright law has already been worked out in detail and only remains to be enacted, one might say the expression refers to a specific new discourse entity. (The question is, which would sound better immediately following the sentence above: (a) *they intend to enact a copyright law next year* or (b) *they intend to enact the copyright law next year*?) Statements of this kind and other borderline cases represented a small proportion (less than 5%) of the total and were evaluated on a case-by-case basis.

Test 3: Subject NPs versus direct-object NPs, considering just specific indefinite NPs, and considering just the first 100 such NPs of each type (subject and direct-object) in the WSJ corpus.

Result: Average length of subject NPs = 5.95 words; average length of direct-object NPs = 8.95 words. ( $F(1, 198) = 13.8, p < .0005$ )

A large and significant difference in length between subject and direct-object NPs persists, even when only specific indefinite NPs are considered. This suggests that the greater length of direct-object NPs is not simply an artifact of a length difference between given and new discourse entities.

Another way of testing DLMR 2 is with subordinate “adverbial” clauses—those introduced by a subordinating conjunction such as *when*, *if*, *as*, or *because*. Adverbial clauses may either be placed before a main clause, as in (14) below, or after it, as in (15). In either case, the parent head of the clause is assumed to be the main verb of the sentence.<sup>8</sup> The head of the adverbial clause itself is the subordinating conjunction, which is almost always the first word of the clause.



While “postmodifying” adverbial clauses are right-branching, “premodifying” clauses are left-branching. Thus the same logic applies as with subject and direct-object NPs: Long clauses in a premodifying position create a long dependency between the clause head and the parent head, as in (14) above, and therefore should be avoided to minimize dependency length. The prediction of DLMR 2, then, is that premodifying clauses should be shorter than postmodifying clauses. This prediction has already been tested by Diessel (2005) on a corpus of both speech and written text (fiction and scientific writings); Diessel found an average length of 10.2 for final adverbial clauses and 7.7 for initial adverbial clauses, thus confirming the prediction of

<sup>8</sup> In the Penn Treebank, postmodifying adverbial clauses are at the same level as the main verb and thus are treated as dependents of it. Premodifying adverbial clauses are at the level of the main clause and therefore could be regarded as dependents of the head of the main clause (the finite verb); or they could be treated as “fronted” postmodifying clauses, and thus dependents of the main verb. We assume the latter option; this makes little difference to the argument, however. The main verb and the finite verb of a clause are usually either the same word, or one or two words apart (for example, if the main verb is a participle). Whichever one of these is the parent head of a premodifying clause, it seems clear that the distance from the adverbial clause head to the parent head will be greater for a premodifying clause than that for a postmodifying clause (compare (14) and (15)).

DLMR 2. (Diessel has an alternative explanation for this result, which will be discussed later.) A similar test was done on the WSJ corpus.

Test 4: Premodifying versus postmodifying adverbial clauses

Result: Average length of premodifying clauses = 9.76 words; average length of postmodifying clauses = 11.41 words. ( $F(1,6880) = 113.7, p < .0005$ )

Premodifying adverbial clauses tend to be shorter than postmodifying adverbial clauses, just as in Diessel's study. Thus the prediction of the DLT is confirmed.

Diessel groups adverbial clauses into three semantic categories, based on the initial conjunction: conditional (*if*), causal (*since, as, or because*), or temporal (*when, after, etc.*). He finds that the proportion of premodifying and postmodifying adverbial clauses varies across these three categories: conditional clauses are mainly premodifying, causal clauses are mainly postmodifying, and temporal clauses are fairly evenly divided between the two. One might wonder, then, if the greater length of postmodifying adverbial clauses was an artifact of this difference—perhaps causal clauses tend to be longer than other kinds. To control for this, the test just reported was repeated, but considering only temporal clauses. (Temporal clauses were chosen since they are much more numerous than conditional or causal clauses—as Diessel's study shows—and are also more evenly balanced between postmodifying and premodifying usages.)

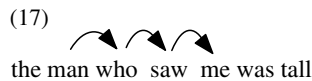
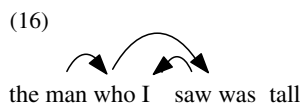
Test 5: Premodifying versus postmodifying temporal adverbial clauses

Result: Average length of premodifying clauses = 9.16 words; average length of postmodifying clauses = 10.83 words. ( $F(1,2552) = 36.3, p < .0005$ )

Even when we consider only temporal adverbial clauses, we find that postmodifying clauses are significantly longer than premodifying clauses. Thus the greater length of postmodifying clauses does not appear to be an artifact of length differences between different semantic types of adverbial clause.

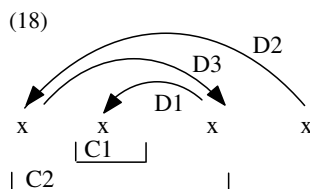
The dependency-lengthening effects of left-branching constituents increase when there are dependencies crossing over the constituent. An example of this is object-extracted relative clauses (or “object RCs”). In an object RC, the RC subject—*I* in (16) below—is left-branching, attaching to the following RC verb, but this dependency is underneath the connection between the relative pronoun and the RC verb. This is in contrast to subject-extracted RCs, as in (17), where there is no such nesting of dependencies.<sup>9</sup>

<sup>9</sup> Regarding the dependency structure of relative clauses, the usual assumption seems to be that the relative pronoun (or complementizer) is dependent on the antecedent noun and the RC verb is then dependent on the relative pronoun; if there is no relative pronoun, the RC verb is directly dependent on the antecedent noun. I follow these assumptions here. (See (Hudson, 1990, p. 389 & 398), and (Gibson, 1998, pp. 20–1), ; this view is also reflected in the head-finding algorithm in Collins, 1999.)

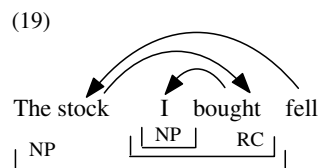


Gibson (1998, 2000) has used this reasoning to explain the greater complexity of object RCs over subject RCs in comprehension. From a production viewpoint, we might expect that the complexity of object RCs would cause them to be used less than subject RCs. Indeed, corpus studies have shown that subject RCs are considerably more common than object RCs in both spoken and written English (Guy & Bayley, 1995; Keenan, 1975). There may well be other reasons for this, however; for one thing, object RCs can only be used with transitive verbs (those that take direct objects), whereas subject RCs can be used with any verb. Other explanations have also been offered for why some kinds of relativization are more common than others (Keenan, 1975). So the greater frequency of subject RCs over object RCs, in itself, is not a very compelling argument for the DLT.

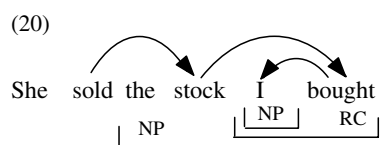
A further prediction that follows from DLMR 2 concerns constituents that are doubly left-branching—that is, a left-branching constituent that is embedded inside another left-branching constituent, such as constituent C1 below.



For constituent C1, DLMR 2 essentially applies twice. Lengthening C1 (assuming again that the head of C1 is near its left end) will lengthen the dependency (D1) between the head of C1 and its parent head; lengthening C1 will also lengthen C2, and hence, the dependency (D2) between C2's head and its parent head. (For that matter, it will also lengthen the right-branching dependency crossing over C1, labeled as D3.) Thus DLMR 2 predicts that there will be extraordinary pressure for C1 to be short. An example in English is subject NPs in relative clauses within subject NPs, such as *I* in the sentence below:



By contrast, RC subjects within direct-object NPs are only crossed by one dependency, so there is less pressure on them to be short.



The prediction is, then, that among subject NPs of relative clauses, those within subject NPs will be shorter than those within direct-object NPs.

Test 6: Subject NPs of relative clauses (RCs), comparing those within subject NPs to those within direct-object NPs

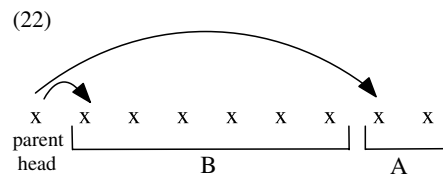
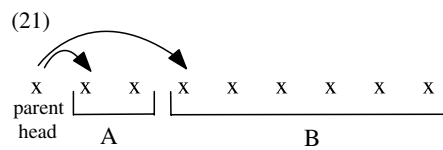
Result: Average length of RC subjects within subject NPs = 1.41; average length of RC subjects within direct-object NPs = 1.61. ( $F(1, 682) = 5.6, p < .05$ )

The prediction is confirmed: Though RC subjects are very short in both subject and object NPs, they are significantly shorter in subject NPs.<sup>10</sup>

### 3. Heads with multiple dependents

So far, we have focused on constructions in which each word just takes a single dependent. We now consider the case where a word has multiple dependents. If we consider just constructions with two dependent constituents, a short one (A) and a long one (B), it can be seen that dependency length will be minimized if the shorter constituent is placed closer to the parent head.

<sup>10</sup> We would also expect object RCs *within* object RCs to be especially rare and difficult to process: e.g. *The stock the man I know sold fell*. In this sentence, the dependency between *I* and *know* is crossed by four other dependencies. These are sometimes known as “center-embedded” constructions (Chomsky & Miller, 1963; Kuno, 1974); the difficulty of processing them is well known, and Gibson (1998, 2000) has argued that it can be accounted for in terms of dependency length.



This leads to a third rule:

DLMR 3. If a word has multiple dependent constituents and there is a choice as to their ordering, the shorter one(s) should be placed closer to the parent head.

In (21) and (22), the two dependent constituents are shown as being internally right-branching, so that the head of each one is at its left end—as is typically the case in prepositional phrases, for example. I have shown elsewhere, however (Temperley, 2005), that the rule just stated does not depend on that. Wherever the internal heads of the constituents are, dependency lengths will be minimized if the shorter one is closer to the parent head.

This rule relates to the well-known phenomenon of syntactic “heaviness” or “end-weight”. In clauses with a direct-object NP and a prepositional phrase, the NP is normally placed first (as in 23a); but if the NP is long, it is often “extraposed”—placed after the PP (as in 23b) (Arnold, Wasow, Losongco, & Ginstrom, 2000; Wasow, 1997, 2002). The same phenomenon is found in dative constructions; if the direct-object NP is long, it tends to be placed after the indirect object (24b) rather than before (24a).

- (23)
- a. The waiter brought [the very expensive red wine we had ordered] [to the table].
  - b. The waiter brought [to the table] [the very expensive red wine we had ordered].

- (24)
- a. Chris gave [a bowl of Mom’s traditional cranberry sauce] [to Terry].
  - b. Chris gave [Terry] [a bowl of Mom’s traditional cranberry sauce].



As Gibson has shown (1998, p. 51), a dependency-length view can explain such phenomena quite straightforwardly; if the long NP in a construction like (23a) is not extraposed, this creates a very long dependency from the verb to the PP (or to the indirect object in 24a). In terms of the current rule system, such phenomena are clearly predicted by DLMR 3.

We can also test DLMR 3 by looking at the ordering of adjunct phrases. In the WSJ corpus, I examined clauses that have two post-verbal adjuncts, where an adjunct is defined as a PP, an ADVP, or an NP-TMP (a temporal noun phrase functioning adverbially, e.g. *he arrived [last week]*). Inspection showed that these three constituent types accounted for the vast majority of non-clausal adjuncts.<sup>11</sup> DLMR 3 predicts that, in such cases, the longer of the two adjuncts will be placed second; thus the mean length of the second adjunct should be greater than the mean length of the first.

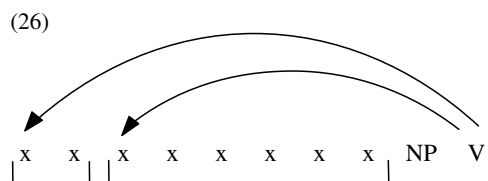
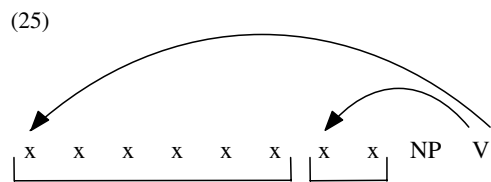
Test 7: Clauses with two postmodifying adjuncts

Result: Average length of first adjunct = 3.04; average length of second adjunct = 5.96. ( $F(1, 5226) = 675.6, p < .0005$ )

The prediction is confirmed; in clauses with two postmodifying adjuncts, the second adjunct tends to be significantly longer than the first.

An interesting comparison here is with premodifying or “fronted” adjuncts—those that precede the subject and verb. In the case of premodifying adjuncts, as with adverbial clauses, the parent head is presumably the main verb of the main clause. Thus in a sentence with two premodifying adjuncts, the second one is closer to the parent head; in this case, according to DLMR 3, dependency length will be minimized if the shorter constituent is placed *second*, as in (25) below, rather than first, as in (26).

<sup>11</sup> Clausal adjuncts (adverbial clauses) were not included; they virtually always seem to follow all other adjuncts, so that this might be regarded as a grammatical rule rather than a matter of syntactic choice. One problem here is that the Penn Treebank does not distinguish adjuncts from arguments. An argument is an element that participates in the event defined by a verb, whereas an adjunct provides additional information; syntactically, arguments are usually defined as elements that are specified by the head, whereas adjuncts are not (Grimshaw, 1990; Pollard & Sag, 1987). ADVPs and NP-TMPs almost always seem to be adjuncts, but PPs are quite often arguments. The problem is that if there were both an ordering difference and a length difference between arguments and adjuncts, this could cause a confound; for example, perhaps arguments tend to come first and also tend to be shorter. The argument/adjunct distinction is not directly represented in the Penn Treebank. However, the treebank documentation (Bies et al., 1995) states that certain kinds of PP types are regarded as “complements” to the VP: those marked with the suffixes -BNF, -CLR, -DAT, -PRD, and -PUT. Thus PPs of these types were considered arguments; all other PPs were considered adjuncts. The same rule was applied in tests 8, 9, 10, and 11 below.



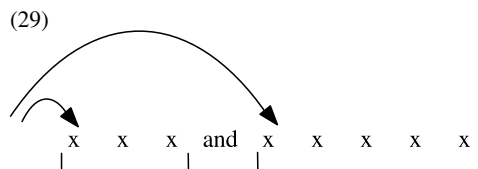
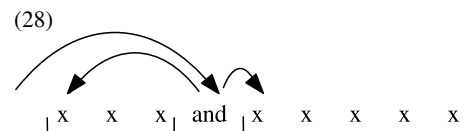
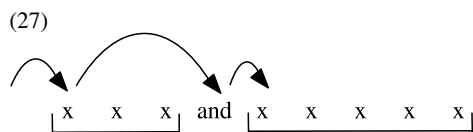
This was examined in the WSJ corpus: clauses with two premodifying adjuncts were examined to compare the lengths of the first and second adjunct. The same adjunct types were considered as in the previous test—PPs, ADVPs, and NP-TMPs.

Test 8: Clauses with two premodifying adjuncts

Result: Average length of first adjunct = 3.15, average length of second adjunct = 3.48. ( $F(1, 1224) = 3.4, p = .06, n.s.$ )

In this case, the prediction—that the first adjunct will be longer than the second—is not confirmed; on average, the second adjunct is slightly longer. However, the difference is small and falls just short of significance. The length ratio of the second adjunct to the first is only 1.10; in the case of postmodifying adjuncts, by contrast, the ratio of second to first is 1.96. The dependency-length view offers an explanation for this difference between premodifying and postmodifying adjuncts. On the other hand, there also appears to be a general preference for “short-long” ordering, even with premodifying adjuncts.

Another test case for DLMR 3 concerns coordinate phrases, in which two constituents (“conjuncts”) are joined by coordinating conjunctions such as *and* or *or*. There is (in many cases at least) a choice as to the ordering of the conjuncts; the question is whether this might be affected by dependency length. I have examined this difficult issue elsewhere (Temperley, 2005) and will only discuss it briefly here. The difficulty lies in determining the dependency structure of a coordinate phrase. At least three proposals have been offered in this regard. Mel’cuk (1987) suggests that the first conjunct of the phrase is the head (connecting to a word outside of the phrase), the conjunction is a dependent of it, and the second conjunct is a dependent of the conjunction, as in (27) below. Munn (1993) proposes that the conjunction is the head, and the two conjuncts are dependents of it, as in (28). Hudson (1990) and Pickering and Barry (1993) argue that both conjuncts act as heads, forming dependency connections with words outside the phrase, as in (29) (under this proposal, the conjunction is essentially left out of the dependency structure).

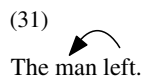
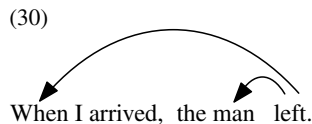


These three proposals offer different predictions regarding dependency length. For right-branching coordinate phrases (in which the entire coordinate phrase is to the right of its external head), all three proposals predict shorter dependency lengths with a short-long ordering of conjuncts. In a study using the WSJ corpus, I found a significant preference for short-long ordering for a variety of right-branching coordinate constructions, including direct-object NPs, prepositional-object NPs, post-verbal PPs, and predicative adjective phrases. For left-branching coordinate phrases, the situation is more complex. The Mel'cuk proposal (27) yields shorter dependency lengths for a short-long ordering, the Hudson proposal (29) favors a long-short ordering, and the Munn proposal (28) expresses no preference. My corpus analysis revealed a short-long preference for left-branching coordinate structures, notably subject NPs. This would seem to favor the Mel'cuk proposal; but the Mel'cuk proposal seems less convincing in other respects. In short, under any syntactic analysis, the dependency-length view predicts a short-long ordering of conjuncts in right-branching coordinate phrases, and this prediction is borne out; the predictions regarding left-branching coordinates are less clear and depend on the syntactic analysis that is assumed.

Our discussion of heads with multiple dependents has focused on cases where there is a choice in the ordering of dependents. But suppose there is no choice? In that case, dependency length will be minimized if the closer dependent is as short as possible.

**DLMR 4.** If a word has multiple dependents and there is no choice as to their ordering, dependents closer to the head should be short.

As a test of this, we can consider subject NPs in main clauses that have premodifying adjuncts (30) versus those that do not (31). (In this case, adjuncts were defined to include adverbial clauses as well as phrasal adjuncts—PP, ADVP, and NP-TMP.)



In a sentence like (30), the main verb *left* has two left-branching dependents (the subject NP *the man* and the premodifying adverbial clause *when I arrived*). There is not really a choice as to their ordering—to place the adverbial clause after the subject NP would seem bizarre (*?The man when I arrived left*). A long subject NP could create an excessively long dependency between the premodifying clause and the main verb; DLMR 4 states that this can be avoided by keeping the subject NP short. In contrast, when there is no premodifying adjunct, as in (31), this pressure for a short subject phrase is not present.

Test 9: Subject NPs in main clauses, with or without premodifying adjuncts.  
 Result: With adjuncts, subject NP average length = 3.17; without adjuncts, subject NP average length = 4.10. ( $F(1, 34498) = 380.2, p < .0005$ )

The prediction is confirmed: subject NPs tend to be shorter when there is a premodifying adjunct.

Adding a premodifying adjunct to a sentence obviously makes it longer; one might wonder if the shorter length of subject NPs in such sentences was due simply to a general avoidance of very long sentences. To examine this possibility, I also examined direct-object NPs in main clauses with and without premodifying adjuncts: [*When I arrived*], *I saw* [*a dog*]. In this case, the premodifying adjunct *when I arrived* and the NP *a dog* are on opposite sides of the parent head *saw*; thus the current theory predicts that the presence of the adjunct will have no effect on the length of the object NP.

Test 10: Direct-object NPs in main clauses, with or without premodifying adjuncts  
 Result: With adjuncts, direct-object NP average length = 7.67; without adjuncts, direct-object NP average length = 7.93. ( $F(1, 11239) = 2.8, p = .09, n.s.$ )

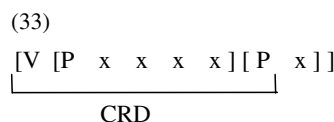
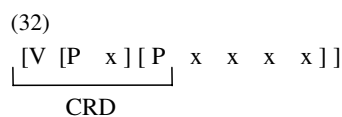
Just as predicted, there is no significant difference in the length of direct-object NPs depending on the presence of a premodifying adjunct. This test seems to rule out the possibility that the shorter length of subject NPs in sentences with premodifying

adjuncts is simply due to a general avoidance of long sentences. If such a factor were operating, it should affect direct-object NPs as well as subject NPs; but the presence of a premodifying adjunct seems to have no effect on the length of direct-object NPs.

Tests 9 and 10 are important in another way, as well. In some of the tests presented here, one might possibly explain the observed effects in terms of discourse or informational factors. For example, with regard to Test 4, one might claim that premodifying and postmodifying adverbial clauses are somehow different in discourse function and that their length difference is due to that. However, it is very difficult to see how the informational status of a subject NP would be affected by the presence of a premodifying adjunct. And if there was such an effect, it is even more difficult to see why it would influence the subject NP but not the direct-object NP.

#### 4. Alternative explanations

It might be argued that some of the findings presented here could be accounted for by principles other than dependency-length minimization. One alternative theory that we should consider is [Hawkins's Early Immediate Constituent \(EIC\) Theory \(1994, 2004\)](#). The EIC theory states that language comprehension will be facilitated if, within each constituent, the heads of the children are all close together—within a short “window”, known as the “constituent recognition domain” (CRD); this is advantageous as it provides the parser with “earlier and more rapid access” to the children of the larger constituent (1994, p. 66). One prediction that follows from the EIC theory is that, in cases where a head is followed by two dependent constituents, the shorter one should be placed closer to the head. In the case of a verb with two adjunct phrases, for example, the shorter adjunct should be placed first, as in (32) below, rather than second, as in (33), since this will minimize the size of the CRD—the portion of the sentence containing the adjunct heads and the parent head.

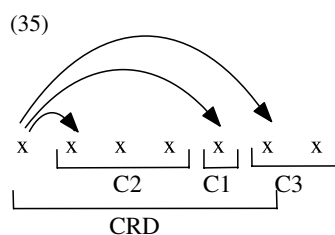
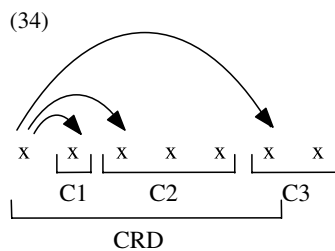


In this way, the EIC theory predicts the finding—reported earlier—that the first adjunct in two-adjunct constructions tends to be shorter than the second.

Another general finding emerging from the tests presented above is that left-branching constituents in English tend to be short. As Hawkins observes, this prediction, too,

follows from the EIC theory: a long left-branching constituent will tend to increase the CRD of the larger constituent. Hawkins suggests that this might account for certain rules of English, such as the fact that phrasal adjectival modifiers are permitted only after a noun (*a book interesting to read*), not before it (*\*an interesting to read book*) (1994, pp. 282–93). In a similar way, the EIC theory might account for the avoidance of long left-branching constituents in situations of syntactic choice. In the case of subject versus object NPs, for example, a long subject phrase will increase the constituent recognition domain of the larger S constituent; by contrast, a long direct-object phrase should not increase the CRD of the VP, as the verb and the head of the object NP will usually remain close together. Diessel (2005) has also invoked the EIC theory as an explanation for the greater length of final versus initial adverbial clauses.

How can we decide between the DLT and the EIC theory? One decisive test concerns heads with three right-branching dependent constituents. In this case, the logic of the EIC theory suggests that the ordering of the first two dependents should make no difference; what matters is the length of the CRD—the distance between the parent head and the head of the last dependent—and that will be the same under either ordering.<sup>12</sup> By contrast, the DLT predicts that the first dependent will tend to be shorter than the second, since this will minimize dependency lengths. In the sentence below, for example, the ordering C1–C2–C3 yields shorter dependencies than the ordering C2–C1–C3; but the CRD is the same size in both cases.



<sup>12</sup> Hawkins does propose a method for predicting ordering preferences in cases where two orderings yield CRDs of the same length. Under this method, the “IC-to-word” ratio (the ratio of constituents to words) is calculated in a left-right manner, adding in one constituent at a time, and then these ratios are averaged (1994, pp. 82–3). Whatever the merits of this metric, it is clearly not ideal to have to posit a special principle for “same-length-CRD” situations; by contrast, the dependency-length view does not require any special treatment of such situations. (A similar point could be made about Hawkins’s handling of head-first versus head-last languages, discussed further below.)

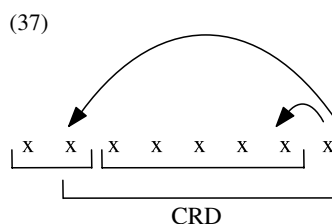
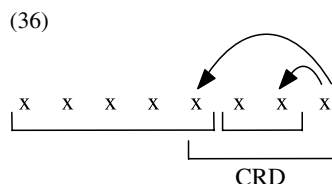
We can test these divergent predictions by looking at clauses with three postmodifying adjuncts, focusing in particular on the first and second adjuncts. (Adjuncts were defined in the same way as in tests 7 and 8 above: PPs, ADVPs, or NP-TMPs.) The DLT predicts that the first will be shorter; the EIC theory predicts no difference in length.

Test 11: In clauses with three postmodifying adjuncts, comparison of the first and second

Result: Average length of first adjunct = 2.98, average length of second adjunct = 3.65. ( $F(1, 790) = 17.2, p < .0005$ )

The first adjunct in a three-adjunct construction tends to be significantly shorter than the second. In this case, then, the DLT appears to fit the evidence better than the EIC theory.

Another interesting source of evidence in this regard comes from head-last languages such as Japanese. Several authors have observed that head-last languages often favor a long-short ordering of constituents. This phenomenon was first studied by Hawkins himself, who cites corpus evidence from Japanese that long-short orderings are indeed preferred—for example, in cases with two NP dependents of a following verb. Other studies have obtained similar findings. An experimental study by Yamashita and Chang (2001) finds a preference for long-short ordering of Japanese subject and object NPs; Yamashita (2002) reports the same pattern in a corpus study of written Japanese.<sup>13</sup> The current theory (specifically DLMR 3) predicts the long-short preference in head-last languages quite straightforwardly: It can be seen that (36) below results in shorter dependencies than (37).



<sup>13</sup> Also relevant here is Hsiao and Gibson's (2003) study of relative clause processing in Chinese. In Chinese, where relative clauses precede their head nouns, object relative clauses have shorter dependency lengths than subject relative clauses, and thus are predicted by the DLT to cause lower processing complexity; an experimental study shows that this is the case.

Hawkins notes that, in a head-last language, a long-short ordering will minimize the length of the CRD (as can be seen in the diagrams above), and argues that it is desirable for this reason. However, this seems to go against the logic of the EIC theory; clearly, the placement of constituent heads towards the end of a larger constituent does not facilitate their “early” identification. Hawkins has proposed that head-last languages feature an entirely different, “bottom-up” parsing strategy, in which the clustering of constituent heads towards the end of a larger constituent is advantageous (1994, pp. 66–7). By contrast, the DLT accounts for the short-long preference in head-first languages and the long-short preference in head-last languages by exactly the same reasoning. Thus the DLT appears to offer a simpler explanation of these phenomena.

A final opportunity for deciding between the DLT and the EIC theory comes from some data from Hawkins (1994) concerning Turkish. In Turkish, a verb-final language, noun phrases may be either head-first or head-last, and there is often a choice as to their ordering. (Hawkins refers to head-first NPs as *mNP* and head-last NPs as *NPm*, and we will do so here as well.) Thus, in a verb phrase containing both an *mNP* and an *NPm*, both *mNP NPm V* and *NPm mNP V* orderings are possible. Fig. 2 shows these two orderings in two situations—where *mNP* is longer than *NPm*, and where *NPm* is longer than *mNP*. The EIC theory makes a straightforward prediction: the *NPm mNP V* ordering should always be preferred. It can be seen that, under the *mNP NPm V* ordering, the CRD is maximally long (extending to the very left edge of the clause), regardless of the length of the two NPs. By contrast, the DLT predicts no general preference for one ordering or the other. The predictions of the two theories are shown in Table 1, along with Hawkins’ data. The data show that overall, the two orderings are roughly equal in preference, with the *mNP NPm V* ordering slightly preferred—contrary to the EIC theory. However, the DLT’s predictions are also not well supported here. The DLT predicts that the shorter constituent should be placed last (closer to the verb); dependency lengths are shorter in Fig. 2a than b, and shorter in Fig. 2d than c. (The EIC theory also predicts that the advantage of the *NPm mNP V* ordering will be especially great when it reflects “long-short” ordering, as in Fig. 2d.) In fact, the data reveal a consistent preference for “short-long” ordering, contrary to both theories. We should note, however, that the body of data available here is very small (87 cases). While this particular test is inconclusive, it points to a possible way of testing the DLT against the EIC theory, given further data.

In this discussion, we have explored three ways of comparing the DLT and the EIC theory. In cases where a verb takes three adjunct phrases, the DLT predicts the observed short-long preference for the first two adjuncts, while the EIC does not. Regarding long-short preferences in head-last languages, the DLT accounts for this phenomenon under the same explanation as short-long ordering in head-first languages, whereas the EIC seems to require rather different explanations for the two cases. In the case of “mixed-branching” constructions in Turkish, while the predictions of the two theories differ, neither one accounts very well for the facts, though only a small data set is available. On balance, then, one could argue that these comparisons slightly favor the DLT, but the issue is certainly far from resolved.



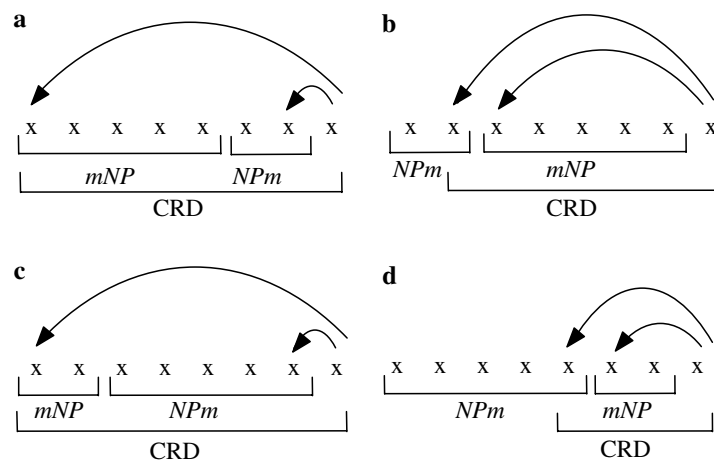


Fig. 2. *mNP NPm V* versus *NPm mNP V* in Turkish, with long *mNP*/short *NPm* (above) and short *mNP*/long *NPm* (below).

Table 1

*mNP NPm V* versus *NPm mNP V* in Turkish, with EIC predictions, DLT predictions, and corpus data

Constituent lengths and ordering	CRD	EIC prediction (* = favored)	Total dependency length	DLT prediction (* = favored)	Number of occurrences (from Hawkins (1994))
<i>mNP</i> = 5, <i>NPm</i> = 2					
<i>mNP NPm</i> (Fig. 2a)	8		8	*	1
<i>NPm mNP</i> (Fig. 2b)	7	*	11		7
<i>mNP</i> = 2, <i>mNP</i> = 5					
<i>mNP NPm</i> (Fig. 2c)	8		8		35
<i>NPm mNP</i> (Fig. 2d)	4	*	4	*	28
Overall					
<i>mNP NPm</i>				No prediction	45
<i>NPm mNP</i>		*			42

Another possible explanation for some of the phenomena observed here lies in an idea suggested by Arnold et al. (2000) (see also Wasow, 2002). Arnold et al. propose that production constraints may lead to a general preference for short-long ordering: “When formulation is difficult, choices in constituent ordering allow speakers to postpone the long, difficult constituent while they utter the shorter, easier one” (2000, p. 32). Several of the results presented above could be explained in terms of a simple “short-first” principle—notably the short-long ordering of adjuncts and coordinate conjuncts. This idea is of particular interest with regard to Test 8, concerning the ordering of pre-modifying adjuncts. The dependency-length view predicts a long-short ordering here, but this was not found; rather, a short-long preference was found (though it is small and statistically non-significant), as would be predicted by the short-first principle. A short-long preference is also found with left-branching coordinate phrases, and with “mixed-

branching” constructions in Turkish; both of these findings contradict the predictions of the DLT, but are well accommodated by a “short-first” view.

While it accounts for many phenomena, the short-first rule cannot explain all of the findings reported here. In particular, it does not appear to predict the general tendency towards short left-branching constituents, found most notably in the case of subject versus object phrases; nor does it predict the shorter length of subject phrases in cases with premodifying adjuncts versus those without. (The short-first principle says that, given a choice, shorter constituents should be placed first; but this is not the same as saying that constituents early in the sentence should be made short.) The short-first principle also has problems with some cross-linguistic data, discussed above, reflecting a long-short preference in head-final languages. Moreover, the suggestion of a short-first principle by Arnold et al. was made with regard to speech, and its plausibility with regard to written language is less clear (we return to this point in the next section). It appears, however, that all of the phenomena observed in this paper can be accounted for either by dependency-length minimization or by the short-first principle. Perhaps the most satisfactory model will prove to be one that combines these two ideas.

## 5. Further Issues

The tests presented here provide compelling evidence that English writers favor syntactic structures with shorter dependencies. The theory is supported by a number of phenomena: the greater length of subject NPs in inverted versus uninverted quotation constructions, the greater length of direct-object versus subject noun phrases, the greater length of postmodifying versus premodifying adverbial clauses, the greater length of relative-clause subjects within direct-object NPs versus subject NPs, the tendency towards “short-long” ordering of postmodifying adjuncts, and the shorter length of subject NPs (but not direct-object NPs) in clauses with premodifying adjuncts versus those without. These findings accord well with [Gibson’s proposal \(1998, 2000\)](#) that increased dependency length causes increased syntactic complexity. However, they also raise several further questions.

One might wonder, first of all, whether the phenomena observed in the WSJ corpus are truly representative of written English generally. (Perhaps the Wall Street Journal has a “style sheet” that recommends short-long ordering of adjunct phrases, for example.) To explore this, all of the tests described above were run on another portion of the Penn Treebank, the Brown corpus. Like the WSJ corpus, the Brown corpus is a large corpus of written English (about 400,000 words), but it is drawn from a broad range of texts—newspapers, magazines, government documents, fiction, and other things.<sup>14</sup> [Table 2](#) shows results of the 11 tests reported in this paper on both the WSJ corpus (as reported above) and the Brown corpus. The results from the Brown corpus will not be analyzed in detail here; suffice it to say that most of the

---

<sup>14</sup> The original Brown corpus ([Francis & Kucera, 1964](#)) contained over 1 million words, but only part of the corpus was syntactically annotated and included in the Penn Treebank.

Table 2  
Tests of constituent length on the WSJ Corpus and the Brown Corpus

Test	WSJ Corpus		Brown Corpus	
	No. of cases	Average length	No. of cases	Average length
1. S–V inversion in quotations				
Not inverted	2305	2.16	540	1.17
Inverted	1768	9.47	49	2.90
2. Subject-object NPs				
Subjects	80444	3.13	39550	1.89
Objects	44092	5.80	18081	3.84
3. Specific indefinite subject-object NPs				
Subjects	100	5.95	—	—
Objects	100	8.95	—	—
4. Adverbial clauses				
Postmodifying	4322	11.41	2382	10.10
Premodifying	2560	9.76	1539	9.21
5. Adverbial temporal clauses				
Postmodifying	1783	10.83	1128	9.06
Premodifying	771	9.16	677	8.91
6. Relative clause subjects				
Subject position	240	1.41	154	1.34
Object position	444	1.61	232	1.34
7. Verbs w/two postmodifying adjuncts				
First	2614	3.04	2323	2.32
Second	2614	5.96	2323	5.22
8. Premodifying adjuncts				
First	613	3.15	232	2.44
Second	613	3.48	232	4.22
9. Subject NPs with/without premodifying adjuncts				
With	9214	3.17	3775	2.01
Without	25286	4.10	10157	2.22
10. Object NPs with/without premodifying adjuncts				
With	3391	7.67	1399	5.09
Without	7850	7.93	3614	4.54
11. Verbs w/three postmodifying adjuncts				
First	396	2.98	256	1.93
Second	396	3.65	256	2.96

basic patterns found in the WSJ data are also reflected in the Brown data. In the Brown corpus as in the WSJ corpus, subject NPs in quotation constructions are longer in inverted than in non-inverted constructions (test 1); object NPs are much longer than subject NPs (test 2); postmodifying adverbial clauses are longer than premodifying ones (tests 4 and 5); the second of two postmodifying adjuncts is longer than the first (tests 7 and 11); and subject NPs, but not object NPs, are shorter when a premodifying adjunct is present (tests 9 and 10). The short-long pattern

for premodifying adjuncts observed in the WSJ corpus—contrary to the predictions of the current theory—is also found in the Brown corpus, indeed much more strongly (test 8). One interesting difference between the two corpora is that in the Brown corpus, unlike the WSJ corpus, subject NPs in relative clauses are not significantly longer in object position than in subject position (test 6). And there are, of course, differences between the two corpora in the magnitude of the various effects. These differences would certainly be interesting to investigate. Broadly speaking, however, the findings from the Brown corpus seem to support the current thesis that syntactic choices in English reflect a preference for structures with shorter dependencies.

If it is accepted that syntactic choices are affected by dependency-length minimization, one might ask why this occurs. One possibility is that it is due to production constraints: constructions with shorter dependencies are less taxing for the producer. Another possibility is that it is a mechanism to facilitate comprehension. Some previous work on syntactic choices has focused on production constraints. In one recent study, Ferreira and Dell (2000) explain the use of the optional *that* with embedded clauses as a time-buying strategy to facilitate fluent sentence production; speakers are more likely to use *that* in cases where the subject of the following clause was not recently mentioned and hence is cognitively less accessible. Similarly, as mentioned above, Arnold et al. (2000) argue that heavy-NP shift and dative alternation are—at least in part—strategies to facilitate production; longer constituents are saved for later because they take more time and effort to plan (though Arnold et al. also recognize the possible influence of comprehension factors).<sup>15</sup> Other authors explain patterns of syntactic choice as strategies to facilitate comprehension. Hawkins, whose EIC theory was discussed extensively above, argues that the grouping of constituent heads within a short “window” aids parsing (though Hawkins has suggested that this may be beneficial for production as well [2004, pp. 110–11]). Temperley (2003) argues that the use of optional *that* in object relative clauses may reflect a strategy of ambiguity avoidance, thus facilitating comprehension in another way. Diessel (2005) also emphasizes comprehension factors, arguing that adverbial clauses are placed sentence-initially in cases where they are necessary for understanding the main clause.

In considering whether syntactic choices are due primarily to production or comprehension pressures, it may be important to distinguish between speech and writing. Both Ferreira and Dell (2000) and Arnold et al. (2000) are primarily concerned with spoken language, arguing that syntactic choices facilitate rapid and fluent speech production. In written language, by contrast, the intense time pressures of speech are not present; in that case, it seems less plausible that producers insert an optional *that* in order to maintain fluency, or place a short constituent first to give themselves

<sup>15</sup> Gibson (1998, p. 52) also seems to assume that, to the extent that dependency-length minimization is reflected in production, it is a mechanism to facilitate production rather than comprehension. While many authors have favored “functional” explanations of syntactic choices—viewing them as strategies to facilitate production or comprehension (or both)—this is not the only possibility. An alternative view is offered by Stallings, MacDonald, and O’Seaghdha (1998). These authors explain heavy-NP shift as an emergent feature of a sentence production process involving competition between alternative syntactic structures, mediated by a variety of syntactic, lexical, and semantic constraints.

a few extra seconds to formulate the longer constituent. It seems more likely, rather, that syntactic choices in writing reflect strategies to facilitate comprehension. Hawkins (1994, 2004), Temperley (2003), and Diessel (2005) all focus on phenomena in written corpora, and they offer comprehension-based explanations for these phenomena. Similarly, in the current case, it seems most plausible to attribute dependency-length minimization in written English to comprehension rather than production factors. It is possible, however, that dependency-length minimization holds benefits for production as well; this remains an unexplored issue.

Would we expect to find dependency-length minimization effects in speech? The answer to this question is by no means obvious. In a way, the idea that speakers consider dependency length in constructing sentences may seem implausible. Calculating the dependency lengths involved in two different constructions is a fairly complex process, and it might seem that there is hardly time for it in the course of speech production. On the other hand, it may be that strategies for dependency-length minimization such as those discussed above—the avoidance of long subject NPs, for example—are acquired by speakers as general habits, so that elaborate length calculations are not required on every occasion. At least one phenomenon predicted by the DLT *has* been observed in speech—the “end-weight” phenomenon, reflected in such patterns as heavy-NP shift (Arnold et al., 2000; Stallings et al., 1998); but as noted earlier, other explanations have been put forth for this as well. In any case, an investigation of dependency-length effects in speech data would certainly be of interest.

A further question concerns the process whereby dependency-length minimization occurs. Two basic, extreme possibilities are a “bottom-up” process and a “top-down” process. In the top-down process, the higher-level structure of the sentence is decided first, and lower-level constituents are then constructed in such a way as to minimize dependency lengths. In the bottom-up process, lower-level constituents are constructed first and are then combined into a larger structure in such a way as to minimize dependency lengths. In some respects, a bottom-up view seems more compelling. With regard to phenomena such as heavy-NP shift, for example, it is difficult to see how the higher-level syntactic structure could be created before the lower-level constituents; if the direct-object NP had not yet been created, the system would not know that it was long and needed to be extraposed. A bottom-up account is also possible with regard to the subject-object length difference: the writer might construct two NPs, a logical subject and logical object, and then choose an active or passive construction partly on the basis of which NP was longer. In this case, however, a top-down view also seems possible: the writer might decide on an active S-V-O construction, but then take dependency length into account in constructing the constituents. For example—as suggested earlier—dependency length might affect decisions as to whether or not to add elaborative information to an NP; the writer might be more likely to do so if the NP in question was the direct object rather than the subject, thus respecting the preference for shorter subject NPs.

This discussion has pointed to several areas that invite further study. Further investigation of cross-linguistic corpus data—building on Hawkins’ pioneering

work—could be of great value, particularly in deciding between the dependency-length theory and other proposals such as the EIC theory and the “short-first” principle. The examination of dependency-length phenomena in speech would also be of interest; among other things, this might shed light on the reasons for dependency-length minimization (the “production versus comprehension” issue discussed above). Further work is needed, too, to determine whether the phenomena discussed here are truly due to dependency-length minimization as opposed to other factors—for example, discourse factors—that may not yet have been considered. All of these issues await investigation as we seek to gain a greater understanding of the role of dependency length in syntactic processing and complexity.

## Appendix A

This appendix shows how the constituent types in the tests presented above were defined in terms of Penn Treebank notation. The underlined constituent is the one whose length was actually measured. There may also be preceding and/or following context which is part of the definition of the constituent.

In the notation below, “\*” refers to any string of alphanumeric characters. [A|B] means “either A or B”. “...” refers to zero or more complete constituents (terminal or non-terminal) of any type.

In these tests, all “empty categories” (constituents containing no overt lexical items) in the Treebank data were ignored; see note 6.

### Test 1

Subject NP in uninverted quotation expression:

([S S-\*] ... (S-TPC\* ...) ... (NP-SBJ\* ...) ...)

Subject NP in inverted quotation expression:

(SINV\* ... (S-TPC\* ...) ... (NP-SBJ\* ...) ...)

### Test 2

Subject NP:

([S S-\*] ... (NP-SBJ\* ...) ...)

Direct-object NP:

(VP ... (NP ...) ...)

### Test 3

The same as Test 2, with the added requirement that constituent being measured (NP or NP-SBJ\*) must be a specific indefinite NP (see note 7 for explanation).

### Test 4

Postmodifying adverbial clause:

(VP ... ([SBAR-TMP|SBAR-ADV|SBAR-PRP] ...) ...)

Premodifying adverbial clause:

([S S-\*] ... ([SBAR-TMP|SBAR-ADV|SBAR-PRP] ...) ...)

## Test 5

Postmodifying temporal adverbial clause:

(VP ... (SBAR-TMP ...) ...)

Premodifying temporal adverbial clause:

([S| S-\*] ... (SBAR-TMP ...) ...)

## Test 6

RC subject NP within subject NP:

(NP-SBJ\* ... (SBAR\* ... ([S| S-\*] ... (NP-SBJ\* ...) ...)) ...)

RC subject NP within direct-object NP:

(VP ... (NP ... (SBAR\* ... ([S| S-\*] ... (NP-SBJ\* ...) ...)) ...)) ...)

## Test 7

Clause with two postmodifying adjunct phrases.

First adjunct:

(VP (V\* ...) ([PP\*| ADVP\*| NP-TMP]...) ([PP\*| ADVP\*| NP-TMP] ...))

Second adjunct:

(VP (V\* ...) ([PP\*| ADVP\*| NP-TMP] ...) ([PP\*| ADVP\*| NP-TMP]...))

## Test 8

Clause with two premodifying adjunct phrases.

First adjunct:

([S| S-\*] ([PP\*| ADVP\*| NP-TMP] ...) ([PP\*| ADVP\*| NP-TMP] ...) (NP-SBJ\* ...)) ...)

Second adjunct:

([S| S-\*] ([PP\*| ADVP\*| NP-TMP] ...) ([PP\*| ADVP\*| NP-TMP] ...) (NP-SBJ\* ...)) ...)

## Test 9

Subject NP in main clause with premodifying adjunct:

(TOP ([S| S-\*] ... ([SBAR\*| PP\*| ADVP\*| NP-TMP]...) ... (NP-SBJ\* ...) ...)) ...)

Subject NP in main clause with no premodifying adjunct:

(TOP ([S| S-\*] (NP-SBJ\* ...) ...)) ...)

## Test 10

Direct-object NP in main clause with premodifying adjunct:

(TOP ([S| S-\*] ... ([SBAR\*| PP\*| ADVP\*| NP-TMP]...) ... (NP-SBJ\* ...) ... (VP ... (NP ...) ...)) ...)

Direct-object NP in main clause with no premodifying adjunct:

(TOP ([S| S-\*] ... (NP-SBJ\* ...) ... (VP ... (NP ...) ...)) ...)

Note: In these two definitions, VPs may be recursive—the object NP may be with any number of VPs.

## Test 11

Clause with three postmodifying adjuncts.

First adjunct:

(VP (V\* ...) ([[PP\*|ADVP\*|NP-TMP] ...]) ([PP\*|ADVP\*|NP-TMP] ...)([PP\*|ADVP\*|NP-TMP] ...))

Second adjunct:

(VP (V\* ...) ([PP\*|ADVP\*|NP-TMP] ...) ([[PP\*|ADVP\*|NP-TMP] ...]) ([PP\*|ADVP\*|NP-TMP] ...))

Note: In tests 7, 8, 9, 10, and 11, PPs carrying the subscripts -BNF, -CLR, -CTV, -PRD, and -PUT were excluded.

## References

- Abbott, B. (2004). Definiteness and indefiniteness. In L. R. Horn & G. Ward (Eds.), *Handbook of Pragmatics* (pp. 122–149). Oxford: Blackwell.
- Abney, S. 1987. The English noun phrase in its sentential aspect. Ph.D. dissertation, MIT.
- Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering. *Language*, 76, 28–55.
- Behagel, O. (1932). *Deutsche Syntax: Eine Geschichtliche Darstellung, Wortstellung, Periodenbau*. (Vol. IV) Heidelberg: Carl Winters.
- Bies, A., Ferguson, M., Katz, K., MacIntyre, R., Tredinnick, V., Kim, G., Marcinkiewicz, M. A., & Schasberger, B. (1995). *Bracketing Guidelines for Treebank II Style Penn Treebank Project*. Technical report, University of Pennsylvania.
- Branigan, H. P., McLean, J. F., & Reeve, H. (2003). Something old, something new: addressee knowledge and the given-new contract. In *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*, Boston.
- Bresnan, J. (1982). *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.
- Chomsky, N. (1988). *Language and problems of knowledge: the Managua lectures*. Cambridge, Mass: MIT Press.
- Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York: Wiley.
- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Norwood, NJ: Ablex.
- Collins, M. (1999). *Head-driven statistical models for natural language parsing*. Ph.D. Dissertation, University of Pennsylvania.
- Diesing, M. (1992). *Indefinites*. Cambridge, MA: MIT Press.
- Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, 43, 449–470.
- Dik, S. C. (1989). *The theory of functional grammar. Part I: The structure of the clause*. Dordrecht: Foris Publications.
- Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40, 296–340.
- Francis, W., & Kucera, H. (1964). *Manual of information to accompany a standard corpus of present-day edited American English, for use with digital computers*. Providence, RI: Department of Linguistics, Brown University.
- Frazier, L. (1985). Syntactic complexity. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 129–189). Cambridge, UK: Cambridge University Press.
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 1–76.



- Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.
- Gibson, E., & Pearlmutter, N. (1994). A corpus-based analysis of psycholinguistic constraints on PP attachment. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 181–198). Hillsdale, NJ: Erlbaum.
- Grimshaw, J. (1990). *Argument Structure*. Cambridge, MA: MIT Press.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.
- Guy, G. R., & Bayley, R. (1995). On the choice of relative pronouns in English. *American Speech*, 70, 148–162.
- Hawkins, J. A. (1983). *Word order universals*. New York: Academic Press.
- Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press.
- Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford, UK: Oxford University Press.
- Heim, I. (1983). File change semantics and the familiarity theory of definiteness. In R. Bauerle, C. Schwarze, & A. vonStechow (Eds.), *Meaning, use and the interpretation of language* (pp. 164–189). Berlin: Walter de Gruyter.
- Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, 90, 3–27.
- Hudson, R. (1990). *English word grammar*. Oxford, UK: Blackwell.
- Keenan, E. L. (1975). Variation in universal grammar. In R. Fasold & R. Shuy (Eds.), *Analyzing variation in English*. Washington: Georgetown University Press.
- King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30, 580–602.
- Kuno, S. (1974). The position of relative clauses and conjunctions. *Linguistic Inquiry*, 5, 117–136.
- Marcus, M., Kim, G., Marcinkiewicz, M. A., MacIntyre, R., Ferguson, M., Katz, K., & Schasberger, B. (1994). The Penn treebank: annotating predicate argument structure. In *Proceedings of the ARPA Human Language Technology Workshop*.
- Mel'cuk, I. A. (1987). *Dependency syntax: Theory and practice*. Albany: State University of New York Press.
- Munn, A. (1993). Topics in the syntax and semantics of coordinate structures. Ph.D. Dissertation, University of Maryland.
- Oehrle, R. T., Bach, E., & Wheeler, D. (1988). *Categorial grammars and natural language structures*. Dordrecht: Reidel.
- Partee, B. H. (1972). Opacity, coreference, and pronouns. In D. Davidson & G. Harman (Eds.), *Semantics of natural language* (pp. 415–441). Dordrecht: Reidel.
- Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, 6, 229–259.
- Pickering, M., & Barry, G. (1993). Dependency categorial grammar and coordination. *Linguistics*, 31, 855–902.
- Pollard, C., & Sag, I. A. (1987). *Information-based syntax and semantics*. Stanford, CA: Center for the Study of Language and Information.
- Radford, A. (1997). *Syntactic theory and the structure of English*. Cambridge, UK: Cambridge University Press.
- Rijkhoff, J. (1990). Explaining word order in the noun phrase. *Linguistics*, 28, 5–42.
- Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language*, 39, 392–417.
- Temperley, D. (2003). Ambiguity avoidance in English relative clauses. *Language*, 79, 464–484.
- Temperley, D. (2005). The coordinate structure of dependency phrases: a corpus approach. *Journal of Psycholinguistic Research*, 34, 577–601.

- Thornton, R., MacDonald, M. C., & Arnold, J. E. (2000). The concomitant effects of phrase length and informational content in sentence comprehension. *Journal of Psycholinguistic Research*, 29, 195–203.
- Vennemann, T. J. (1974). Topics, subjects and word order: from SXV to SVX via TVX. In J. M. Anderson & C. Jones (Eds.), *Historical Linguistics I (Proceedings of the First International Conference on Historical Linguistics)*. New York: American Elsevier Publishing Company.
- Wasow, T. (1997). End-weight from the speaker's perspective. *Journal of Psycholinguistic Research*, 26, 347–361.
- Wasow, T. (2002). *Postverbal behavior*. Stanford, CA: CSLI Publications.
- Yamashita, H. (2002). Scrambled sentences in Japanese: linguistic properties and motivations for production. *Text*, 22, 597–633.
- Yamashita, H., & Chang, F. (2001). “Long before short” preference in the production of a head-final language. *Cognition*, 81, B45–B55.