Journal of Psycholinguistic Research, Vol. 34, No. 6, November 2005 (© 2005)
DOI: 10.1007/s10936-005-9165-2

The Dependency Structure of Coordinate Phrases: A Corpus Approach

David Temperley[1]

*Hudson (1990) proposes that each conjunct in a coordinate phrase forms dependency relations with heads or dependents outside the coordinate phrase (the "multi-head" view). This proposal is tested through corpus analysis of Wall Street Journal text. For right-branching constituents (such as direct-object NPs), a short-long preference for conjunct ordering is observed; this is predicted by the multi-head view, under the assumption that structures resulting in shorter dependencies are preferred. A short-long preference is also observed for left-branching constituents (such as subject NPs), which is less obviously accommodated by the multi-head view but not incompatible with it. The repetition of determiners was also examined (*the dog and cat *versus* the dog and the cat*), and a stronger preference was found for repetition with singular count nouns as opposed to mass or plural nouns; this accords well with the multi-head view, under the reasoning that single-determiner constructions require crossing dependencies with count nouns but not with plural or mass nouns.*

KEY WORDS: coordination; corpus analysis; dependency grammar; syntactic complexity.

INTRODUCTION

The concept of dependencies—asymmetrical syntactic relations between pairs of words in a sentence—is of central importance in theoretical linguistics, playing a role in almost every major syntactic theory (Bresnan, 1982; Pollard & Sag, 1987; Mel'cuk, 1987; Oehrle, Bach & Wheeler 1988; Dik, 1989; Hudson, 1990; Radford, 1997). In recent years, the concept has also proven to have great value and relevance in psycholinguistics. In particular, the work of Gibson (1998, 2000) has shown that a dependency-based view of syntax sheds light a variety of psycholinguistic phenomena—the central idea being that syntactic structures containing longer dependencies are more complex. This principle accounts for numerous

[1] Eastman School of Music, 26 Gibbs St., Rochester, NY 14604, USA. e-mail: temperley@esm.rochester.edu

observed effects in language comprehension and production, as will be described in greater detail below.

Given the explanatory power of dependencies, it is important to determine what the dependency structures of a language actually are. In many cases, this is uncontroversial. In the case of prepositional phrases in English, for example, it is generally agreed that the preposition is the head of the phrase (with the object noun phrase as its dependent) and is then a dependent of the preceding word (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.[2] In some cases, however, the dependency structure of syntactic constructions is less certain. A case in point is coordinate phrases—phrases involving two or more elements (such as noun phrases or verb phrases) joined together with a coordinating conjunction. The dependency structure of coordinates has received relatively little attention, either in linguistics or in psycholinguistics. However, one proposal has been embraced by several authors—what I will call the "multi-head" proposal. Under this proposal, the conjuncts within the coordinate phrase all serve—in effect—as heads, each one making one or more dependency connections to the rest of the sentence.

In this paper I will attempt to validate the multi-head proposal empirically. My methodology relies on a corpus approach—statistical analysis of naturally-occurring written language—and depends on two fundamental premises, both of them quite well-established. (1) In situations of syntactic choice—where there is more than one way of expressing something—people tend to use the construction that is syntactically less complex or computationally demanding. (2) The complexity of a sentence increases with the length of the dependencies it contains. As noted above, this principle has shown great value in explaining facts of language comprehension and production. (Complexity may be affected by other aspects of dependency structure as well, as I will explain.) Given these

[2] Noun phrases are a somewhat controversial case. While many in linguistics and psycholinguistics have assumed that NPs are headed by their main nouns, with determiners and modifiers as dependents (Bresnan, 1982; Pollard & Sag, 1987; Mel'cuk, 1987; Gibson, 1998), recent theory in the GB/minimalist vein assumes the determiner of an NP as the head (Abney, 1987; Radford, 1997), as does Hudson's Word Grammar (1990). The current study offers new evidence on this issue, as will be explained. A further general complication is that constituents are sometimes assumed to be headed by null elements, i.e., empty categories. The allowance of empty categories raises many problems in terms of the location of heads, calculation of dependency lengths, and evaluation of complexity; given the lack of consensus as to the exact nature of empty categories or even their existence (Pickering & Barry, 1991), we will not attempt to accommodate them here.

two premises, a proposal regarding dependency structures may be taken to imply predictions about the complexity of sentences, and in turn, about language use: in situations of syntactic choice, the constructions that are predicted to be more complex should be used less often. We consider two kinds of syntactic choice in particular: (1) ordering of conjuncts within the coordinate phrase and (2) the repetition of determiners in conjoined noun phrases.

THE MULTI-HEAD THEORY OF COORDINATE DEPENDENCIES

Hudson (1990) offers a proposal for the dependency structure of coordinate phrases. Under this proposal, the words within a conjunct form dependency relations with one another in the normal way. The head of each conjunct then makes connections to words outside the coordinate phrase. (Actually, each conjunct may have more than one head, as will be described below.) This permits a succinct statement of the rules regarding well-formed coordinate structures: Any word which is outside a coordinate phrase and in dependency relation D to the head of a conjunct must also be in relation D to the head of every other conjunct. Thus the sentence *Fred peeled and ate it*—with the verbs *peeled* and *ate* as the two conjuncts—is valid because both verbs have a subject relation with *Fred* and an object relation with *it*. *Fred peeled and yawned it* is incorrect, because *it* is in an object relation with *peeled* but not with *yawned*. Under this framework, the conjunction itself makes no dependency connections but is treated in a special way. The "multi-head" approach to coordinate phrases is also assumed in Sleator and Temperley's Link Grammar (1993), and endorsed, though with some modifications, by Pickering and Barry (1993).

One of the virtues of the multi-head model is that it can allow for cases in which conjuncts are not "phrases" in conventional terms. Each conjunct may consist of more than one phrase, each one having its own head; but the rule above still applies, in that each conjunct must make parallel connections to the rest of the sentence. In a sentence like *I drank coffee at eleven and tea about four*, each conjunct consists of a noun phrase and a prepositional phrase, each one connecting to the previous verb: thus the sentence is correctly allowed. However, the multi-head proposal also violates some normal assumptions of dependency grammar. Most formulations of dependency grammar have assumed that each word must have exactly one head (Gaifman, 1965; Mel'cuk, 1987). Under the multi-head theory, however, multiple conjunct-heads may all serve as heads for a dependent outside the coordinate phrase. In the sentence below, the two conjoined verbs both function as heads for the subject *Fred* (and also for the direct object *it*).

(1)

Second, it is generally assumed that dependencies may not cross—a rule sometimes known as "projectivity" (Gaifman, 1965; Mel'cuk, 1987). Under the multi-head theory of coordination, however, violations of projectivity occur frequently; dependencies will cross any time the conjuncts make more than one connection to the rest of the sentence, as in (1) above. It has been observed that crossing dependencies are rare cross-linguistically (Steedman, 1985), though there are well-known examples in certain languages such as Dutch (Bresnan, Kaplan, Peters, & Zaenen, 1982; Joshi, 1990); in dependency grammars of English, crossing dependencies are generally disallowed, or allowed only under highly constrained circumstances (Hudson, 1990; Sleator & Temperley, 1993; Collins, 1996). (The fact that the multi-head proposal violates projectivity and the one-head-per-dependent rule is not discussed either by Hudson or by Pickering and Barry; however, Hudson does allow other exceptions to these rules.)

Because the multi-head proposal conflicts with these general assumptions about dependencies in English, it is of particular interest to examine it empirically. We now consider how well the multi-head model can account for syntactic choices in the use of coordinate phrases.

ORDERING OF CONJUNCTS IN COORDINATE PHRASES

One situation of syntactic choice in coordinate phrases concerns the ordering of conjoined elements. In principle, language users have a choice as to how the conjuncts in a coordinate structure are ordered; any ordering is generally syntactically correct. Although such choices could be affected by many factors (as discussed below), it seems possible that they might be determined at least in part by syntactic complexity, and thus, by considerations of dependency length. It is of interest to consider whether ordering preferences in coordinate phrases accord with the predictions of the multi-head theory.

The idea that the complexity of syntactic constructions might be related to the distance between heads and dependents has been discussed by several authors (Frazier, 1985; Rijkhoff, 1990; Hawkins, 1994), but assumes an especially important role in Gibson's dependency locality theory (DLT) (1998, 2000). Gibson suggests that syntactic complexity can largely 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. 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.[3] This idea accounts for a number of phenomena in comprehension, such as the greater complexity of object as opposed to subject relative clauses (King & Just, 1991). In both subject and object relatives, the verb of the relative clause must be connected with the preceding relative pronoun, but in the object case the two words are separated by the relative clause subject, yielding a higher integration cost. The theory also relates to phenomena in ambiguity resolution; in main-verb/reduced-relative ambiguities (Gibson, 1998) and prepositional-phrase attachment decisions (Gibson & Pearlmutter, 1994; Thornton, MacDonald & Arnold, 2000), the preferred interpretation reflects a preference for shorter dependencies. Most relevant to the current study, Gibson puts forth the DLT to account for syntactic choices such as "heaviness" effects. It has been widely observed that, when a verb has two arguments (such as an NP and a PP), there is a preference to put the shorter argument first (Bever, 1970; Hawkins, 1994; Arnold, Wasow, Losongco & Ginstrom, 2000). Normally, a direct-object NP occurs before a PP—(2a) below is much more likely than (2b). But if the NP is especially long, it may be extraposed, as in (2c):

(2a) She sold [NP the ring] [PP for five dollars]
(2b) ? She sold [PP for five dollars] [NP the ring]
(2c) She sold [PP for five dollars] [NP the beautiful diamond ring her grandmother had given her when she was twelve]

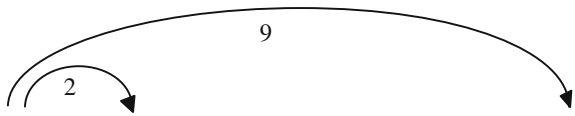
This is predicted straightforwardly by the DLT; both arguments make a dependency connection to the preceding verb, and the total length of these dependencies is minimized if the shorter phrase is placed first.

Returning to the case of coordinate constructions, the multi-head theory predicts that the dependency length—and hence the complexity—of such constructions will be affected by the ordering of the conjuncts. Consider (3) below, in which the direct object NP is a coordinate phrase with two conjuncts; one is 2 words long (*some flowers*) and the other is 6 words long (*a bottle of very expensive wine*). (In this study we will focus on coordinates with two conjuncts, though other cases will be considered briefly in

[3] Gibson measures the length of dependencies not in sheer number of words spanned, but in "discourse referents" spanned. However, he seems to regard this as conjectural, and acknowledges that "[p]rocessing all words probably causes some integration cost increment" (1998, p. 12).

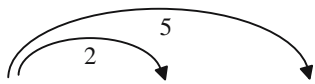
the final section.) Both conjuncts make a single connection to an external head—the "matrix head"—which is immediately to the left of the coordinate phrase. Consider just the lengths (in words spanned) of the two dependencies connecting the conjunct-heads to the matrix head, indicated by the numbers in the diagram below. (Each conjunct-head also makes connections to words within the conjuncts, but this is unaffected by the ordering of the conjuncts; dependencies between words outside the coordinate phrase are also unaffected. We will also assume, following Hudson, that the conjunction itself makes no dependency connections to other words.) If the longer conjunct is first, the total length is $2 + 9 = 11$; if the shorter conjunct is first, the total length is $2 + 5 = 7$. Thus a short-long ordering of the conjuncts results in substantially shorter dependencies.

(3a)



She brought [a bottle of very expensive wine] and [some flowers]

(3b)



She brought [some flowers] and [a bottle of very expensive wine]

This result—that the short-long ordering yields shorter dependency lengths—proves to hold true quite generally with "right-branching" coordinates (where the matrix head is to the left), regardless of the exact length of the conjuncts, the position of the heads within the conjuncts, and the distance from the coordinate phrase to the matrix head (see Fig. 1a). Define $C1$ and $C2$ as the two conjuncts, $L1$ and $L2$ as their respective lengths (in words), $P1$ as the distance in words from the beginning of $C1$ to its head (so $P1 = 0$ if $C1$'s head is its leftmost word), $P2$ as the corresponding value for $C2$, M as the distance from the left end of the entire coordinate phrase to the matrix head, $T1$ as the total length of the two dependencies (connecting **H1** and **H2** to the matrix head) if $C1$ is the first conjunct, and $T2$ as the total length if $C2$ is first. From Figure 1a, it can be seen that

$$T1 = (M + P1) + (M + L1 + 1 + P2)$$

$$T2 = (M + P2) + (M + L2 + 1 + P1)$$

(The "1" in the equations above is the length added by the conjunction itself.) What is of interest here is the "length advantage" of $T2$ over $T1$—the amount that $T2$ is greater than $T1$, or $T2 - T1$. (Of course, greater length is actually not advantageous in this case.)

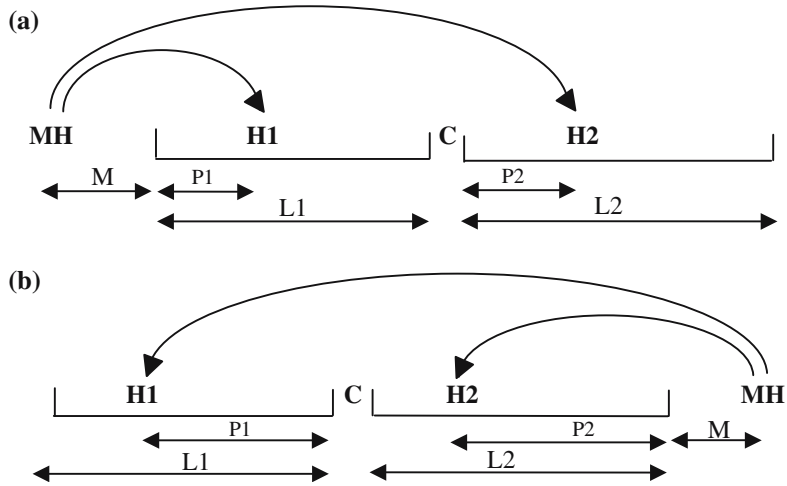


Fig. 1. Dependency lengths in a coordinate phrase with two conjuncts, when the matrix head is to the left (A) or to the right (B). In (A): MH = matrix head, **H1** = head of first conjunct, **H2** = head of second conjunct, M = distance from left end of coordinate phrase to MH, $L1$ = length of first conjunct, $L2$ =length of second conjunct, $P1$ = distance from **H1** to left end of first conjunct, $P2$ = distance from **H2** to left end of second conjunct, C = conjunction. In (B) the symbols are analogous, with MH to the right of the coordinate phrase.

$$\begin{aligned} T2 - T1 &= (2M + P2 + L2 + 1 + P1) - (2M + P1 + L1 + 1 + P2) \\ &= L2 - L1 \end{aligned}$$

This shows that the ordering $(C2, C1)$ will always have a length advantage of $(L2 - L1)$ relative to $(C1, C2)$. If $L2 > L1$, this number will be positive—that is to say, the $(C2, C1)$ ordering will result in greater dependency length, with the magnitude of the difference depending only on the relative length of the two conjuncts.

Now consider the case where the matrix head follows the coordinate phrase (Figure 1b). In this case, the situation in Figure 1a is exactly reversed; the length advantage of $(C2, C1)$ over $(C1, C2)$ is $L1 - L2$, so that the total dependency length will be greater if the shorter constituent is first. The predictions, then, are that a short-long ordering should be preferred when the coordinate phrase follows the head, and long-short ordering preferred when it precedes the head. (We should note, in passing, that a preference for long-short ordering in head-final languages has been noted by Hawkins (1994) and Yamashita and Chang (2001); this would seem to be explained quite naturally in dependency-length terms, though this has not been much explored.)

These predictions were tested empirically using the Penn Treebank (the *Wall Street Journal* portion), a corpus of one million words of naturally occurring text from the 1989 *Wall Street Journal*.[4] The text is tagged with constituent information, permitting data about syntactic constructions to be gathered computationally. A coordinate phrase was defined as a constituent containing exactly three smaller constituents, the second of which was the coordinating conjunction *and* or *or*; the first and third constituents were then assumed to be conjuncts. Data was gathered for different constituent types (explained below). For each constituent type, all cases in the treebank were identified; the length of the first conjunct (in words) in each case was measured, and this was averaged over all cases of that type; the same was done for the second conjunct.

Constituent types were defined not only by a syntactic category (such as "NP" or "VP"), but also by their larger context. The reason for this was that it was important to identify each type as to whether the matrix head was to the right or left (since the multi-head model's predictions differ in these two cases). While dependencies are not explicitly marked in the treebank, in many cases they can be inferred quite reliably from the constituent structures. For example, a PP directly inside a VP was assumed to be a prepositional phrase modifying a preceding verb. Table 1 shows the constituent types tested, and the way each one was defined in terms of Treebank notation. Table 1 also indicates which types were right-branching or left-branching. Most of the common constituent types in English are right-branching, but three left-branching constituent types were tested: subject NPs (which are dependents of the following verb—except in cases of subject-verb inversion), prepositional phrases acting as "openers" (thus connecting to the head of the following clause), and attributive adjective phrases (connecting to a following noun). It should be noted that the definitions are not perfect. For example, a prepositional phrase within a verb phrase may occasionally precede the verb (in which case it is not right-branching). But in each case large portions of the data were inspected by hand, and it seemed that the definition resulted in only a very small number of errors, either false positives (cases incorrectly included in the test) or false negatives (cases incorrectly excluded). Given the limited annota-

[4] Two other corpus studies have examined conjunct length, but did not consider the particular issue investigated here. Gibson, Schütze, and Salomon (1996) examined phrases with two NP conjuncts, and found a greater average length for the second NP (4.06 vs. 3.10), but did not distinguish between different syntactic types of NP. Levy (2002) performed a test of two-NP coordinate phrases, but only comparing NPs in general with sentence-initial NPs; a short-long preference was found for both cases, but was somewhat weaker in the case of sentence-initial NPs.

Table I. Conjunct Length in The Penn Treebank

Constituent type	Treebank definition	Number of cases[a]	Conjunct 1 average length (L1)	Conjunct 2 average length (L2)	L2/L1
Right-Branching Constituents					
Direct object NP	NP(unsubscripted[b]), daughter of VP[c]	785	3.84	5.08	1.32*
Prepositional object NP	NP, daughter of PP	2842	2.63	3.49	1.32*
VP dependent on an auxiliary verb	VP, daughter of VP	806	5.07	6.97	1.38*
PP modifying preceding verb	PP, daughter of VP	162	5.78	7.62	1.32*
Subordinate or embedded clause[d]	S, daughter of SBAR	200	7.43	8.34	1.12
Predicate adjective phrase	ADJP-PRD, daughter of VP	115	1.67	2.77	1.66*
Left-Branching Constituents					
Subject NP	NP-SUBJ, daughter of S[e]	1026	2.49	3.28	1.32*
Attributive adjective phrase	ADJP, daughter of NP, but not at the end	446	1.07	1.11	1.04
PP dependent on following clause	PP, daughter of S (and starting at same point)	7	3.57	7.29	2.04

[a]The number of two-element coordinate phrases of this type observed in the Wall Street Journal portion of the Penn Treebank.
[b]Selecting only unsubscripted NPs excludes NPs that are not direct objects, such as temporal adjuncts like *last week* (tagged as NP-TMP).
[c]X is a daughter of Y if it is contained in Y and immediately subordinate to it.
[d]A clause dependent on an external head to the left, such as a subordinating conjunction, relative pronoun, complementizer, or verb.
[e]Selecting only daughters of S excludes subject NPs in cases of subject-verb inversion (in which case the mother category would be SINV).
*An ANOVA showed that the difference between the $C1$ and $C2$ conjunct lengths was significant, $p < 0.01$.

tion provided in the Penn Treebank, it seemed impossible to devise perfect definitions for these syntactic types.

Table I shows, for each constituent type, the ratio between the average second-conjunct length and the average first-conjunct length; a value greater than one indicates a preference for short-long ordering. Considering first the "right-branching" constituent types, we observe a strong pattern of short-long ordering. This pattern is rather consistent

across the different constituent types—a ratio of between 1.12 and 1.66 is observed in every case. This is in accord with the multi-head model, which predicts a short-long preference for right-branching coordinate structures. However, consider the left-branching cases. Here the model predicts a long-short preference, but this is not observed. The largest body of data here is subject NPs; in this case, a clear short-long preference is found (virtually equal to that observed for right-branching NPs). For attributive adjective phrases, a short-long preference is found, though it is notably less strong than for any other type; in this case, however, the vast majority of conjuncts were only one word long. For "opener" PPs modifying noun phrases, we find a strong short-long preference, though in this case the body of data is too small (only 7 cases) to be given much weight.

Overall, the predictions of the multi-head model are partially borne out. A strong preference for short-long ordering in right-branching constituents is observed, just as predicted by the model. However, while the model predicts a long-short preference in left-branching constituents, a short-long preference is observed in these cases as well. Still, the failure of the multi-head model to predict these cases is not necessarily a fatal problem. I have suggested elsewhere (Temperley, 2003) that syntactic choices tend to be guided by principles of a fairly general nature; rather than choosing the most desirable alternative in every case, language users employ general policies which lead to the best result most of the time.[5] Presumably, most constituents in English are right-branching—the generally "head-initial" character of English is well-attested (Venneman, 1974; Radford, 1997); on balance, therefore, a short-before-long strategy for conjunct ordering would be more beneficial than a long-before-short strategy. Even if we assume that people employ different strategies for different phrase types (NPs, VPs, etc.), it is likely that even for NPs—the most frequent type of left-branching constituent—a consistent short-before-long strategy is more beneficial than a long-before-short one, since right-branching phrases are more common even in this

[5] This claim was originally based on syntactic choices motivated by ambiguity-avoidance. An example is in embedded-clause constructions, like *She said (that) he was coming*. In such cases, people (in both written and spoken language) are more likely to include the optional complementizer *that* when the embedded-clause subject is a full NP rather than a pronoun (Elsness, 1984; Ferreira & Dell, 2000); this can be attributed to ambiguity-avoidance, on the grounds that, if the complementizer were omitted, a full NP could be ambiguous between a direct object and an embedded-clause subject, whereas a pronoun like *he* or *I* is case-specific and causes no such ambiguity. But people make no distinction in this regard between *I* and *you*, even though *I* is case-specific while *you* is not. The use of relative pronouns or complementizers with object relative clauses also seems to reflect such a general strategy (Temperley, 2003).

case. (As Table 1 shows, prepositional–object and direct–object coordinate NPs together are more than three times as numerous as subject NPs; moreover, subject NPs tend to be somewhat shorter than the other two types, so that length considerations may be less important.)

Before proceeding, we should consider whether the preference for short-before-long ordering of conjuncts might be due to other factors besides complexity. Certainly, other considerations might sometimes effect conjunct ordering; these could include temporal order or logical order (*We had steak and ice cream for dinner*), or anaphoric relations between conjuncts (*We saw John and his mother*). Of particular interest is the factor of "newness". Consider sentences (4a) and (4b) below, both from the treebank; in these two cases, the first of the two conjuncts is a previously mentioned item in the discourse, whereas the second item is being mentioned for the first time (the conjuncts are italicized).

(4a) Separately, a federal judge hearing Mr. Hunt's bankruptcy case yesterday turned down a proposed $65.7 million settlement between *Mr. Hunt* and *Minpeco S.A., another major creditor in the case*.

(4b) *Manville* and *a spokeswoman for the trust* said that the two are discussing the proposal but a decision hasn't been made.

Such examples reflect a well-established principle of discourse, sometimes called "given-before-new": within a sentence, entities already present in the discourse tend to be mentioned before new entities (Clark & Haviland, 1977; Arnold *et al.*, 2000). The newness factor is a possible problem, since it may be confounded with length. In cases where one element in a coordinate is old and another is new, it seems likely that the new element would often be longer. Quite often, when a new element is introduced, it is accompanied with descriptive information; this might take the form of an appositive (as in (4a)), or a prepositional phrase or relative clause (as in (4b)). If this is the case, the observed short-long preference might be due partly to newness rather than to syntactic complexity.

Discourse newness is difficult to measure; by many accounts, it is a matter of degree or even a continuum (Clark & Haviland, 1977; Brown & Yule, 1983; Gundel, Hedberg, & Zacharski, 1993). While some entities may be clearly given or clearly new, others have not been explicitly mentioned but are to some degree inferable from the context. (The given-new distinction arises also with other categories besides nominals, but we will only consider nominals here.) It does seem, however, that nominals containing phrasal modifiers of some kind—prepositional phrases, relative clauses, or appositives—are more likely to be new. We may control for newness to some extent, then, by examining coordinate NPs in which either both conjuncts contain phrasal modifiers or neither of them does.

Table II. Conjunct Length in NP Coordinate Phrases in the Penn Treebank[a]

Constituent type	Number of cases	Conjunct 1 average length ($L1$)	Conjunct 2 average length ($L2$)	$L2/L1$
Both conjuncts are base NPs	3139	1.73	1.92	1.11**
Neither conjunct is a base NP	698	7.07	8.25	1.17**
Both conjuncts begin with "a" or "an"	177	5.54	6.59	1.19*

[a]The set of NPs examined here includes all direct-object, prepositional-object, and subject NPs, as defined in Table 1.

*An ANOVA showed that the difference between the C1 and C2 conjunct lengths was significant, $p < .05$.

**An ANOVA showed that the difference between the C1 and C2 conjunct lengths was significant, $p < .0001$.

(NPs containing no phrasal modifiers are known as "base NPs" in treebank terminology.[6]) The data for the treebank is shown in Table 2. It can be seen that, even under these conditions, a length effect persists (though it is somewhat reduced). For coordinates in which neither conjunct has a modifier, the second one is still significantly longer; the same is true when both conjuncts have modifiers. As a further test, cases were examined in which both NPs began with the determiner *a/an*; this is generally assumed to be strongly correlated with newness (Gundel *et al.*, 1993). Here again, the short-long ordering effect persisted, though again somewhat reduced (see Table 2). Thus, while this issue requires further study, it seems unlikely that the length effect in conjunct ordering is due entirely to newness.

Overall, the ordering of conjuncts gives substantial, though not unqualified, support for the multi-head model. The short-long preference for right-branching constituents is as predicted by the model; the short-long preference for left-branching constituents is not directly predicted, but is compatible with the model, by the reasoning that a general policy of short-long ordering would be advantageous given the preponderance of right-branching structures in English. We now examine another kind of syntactic choice in the use of coordinate constructions, which seems to have bearing on the multi-head model, but in quite a different way.

[6] Strictly speaking, a base NP is an NP that does not contain any phrasal constituents of any kind; thus base NPs never contain phrasal modifiers. Inspection of the treebank shows that non-base-NPs contain phrasal post-modifiers in a large majority of cases, though not always; for example, the internal phrase may occasionally be a pre-noun possessive phrase.

DUPLICATION OF DETERMINERS WITH CONJOINED NOUNS

Another common situation of choice in coordinate phrases involves what I will call "duplication". Often, with a coordinate expression, one has a choice of including an element in the coordinate (and thus repeating it), or stating it just once outside the coordinate. For example, one could say *I saw Mary and saw Bill*, or *I saw Mary and Bill*. In many kinds of constructions, duplication hardly ever occurs; *I saw Mary and saw Bill* seems unlikely and unnatural, though not incorrect. In other situations, it is quite common; perhaps the strongest example is the case of determiners modifying nouns, where the determiner is sometimes repeated and sometimes not: One might well say either *The dog and cat were black*, or *The dog and the cat were black*.

Informal inspection of noun phrases with determiners pointed to a surprising pattern. Duplication of the determiner seemed to occur much more often when the nouns involved were singular count nouns, as opposed to plural count nouns or mass nouns. That is, there seemed to be a tendency to favor *the dog and the cat* over *the dog and cat*, but to favor *the dogs and cats* over *the dogs and the cats*, and *the ice and snow* over *the ice and the snow*. One might ask, how is this related to the dependency structure of coordinate phrases? To preview the following argument, a structure like *the dog and cat* involves crossing dependencies according to the multi-head model (assuming that both nouns make a connection to the determiner and also to some other word); crossing dependencies are unusual in English and are presumably difficult to process. When plural or mass nouns are used, as in *the ice and snow*, the second noun does not require a determiner; thus the sentence can be parsed without crossing dependencies (even though the correct analysis *does* involve crossing dependencies), so it should be more easily processed. Thus, if repetition of the determiner is indeed more favored with singular count nouns, this would provide evidence for the multi-head model. We will now follow this line of reasoning in more detail.

The first step is to examine quantitatively the tendency observed above—that repetition of the determiner in coordinate phrases tends to occur more often with singular count nouns than with mass or plural nouns. Again, the Wall Street Journal portion of the Penn Treebank was used. All noun phrases in the corpus of the form "D N C N" were examined (we will call these single-determiner or SD constructions), as well as all phrases of the form "D N C D N" (double-determiner or DD constructions), where D is the determiner *the* or *a/an* or a possessive pronoun, N is any common noun, and C is the conjunction *and* or *or*. (In DD constructions, the first and second D must be the same word. Other

words marked as determiners in the treebank were excluded, as they sometimes seemed to semantically restrict the choice between an SD and DD construction.[7]) It is possible that, in an SD construction in which the second noun was plural or mass, the determiner would not be intended to apply to it; for example, one sentence contained the phrase *the quake and baseball*. The SD phrases were inspected one by one, and all phrases of this type were omitted. Also omitted were cases where the entire phrase seemed to be an idiomatic expression, like *a bed and breakfast*. Another problem concerned cases where the coordinate phrase was followed by a modifying phrase such as a prepositional phrase or relative clause. In such cases, there seems intuitively to be a preference for the SD construction (perhaps because it makes clear that the post-modifier applies to both nouns); one would probably say *the book and CD that we bought* rather than *the book and the CD that we bought* if the relative clause was supposed to apply to both NPs. And it may also be the case that nouns followed by post-modifiers are more likely to be mass nouns than count nouns (informal inspection of the data suggested that this might be true); in this case, any preference for SD constructions with mass nouns might be due to the correlation between mass nouns and post-modifiers rather than to a direct preference. To avoid this possible confound, all NPs with post-modifying phrases were excluded from the test set.

As a first comparison, SD and DD phrases were counted in which both nouns were plural or both were singular. (Nouns are annotated as singular or plural in the treebank; the "singular" category includes both count and mass nouns.) Table 3 shows the results: 40.4% of singular phrases were of the DD type, whereas only 8.8% of plural phrases were ($\chi^2 = 17.4$, $p < 0.0001$). Thus there seems, indeed, to be a significantly stronger preference for omitting the second determiner with plural phrases than with singular phrases. Since singular nouns sometimes require determiners for syntactic correctness (i.e. count nouns), whereas plural nouns never do, this points to a stronger preference for DD constructions with nouns that require determiners.

For comparing count nouns to mass nouns, a more complex procedure was required. Nouns are not marked in the treebank as count or mass, and many nouns allow both count and mass usages. One indication of mass nouns is usage without a determiner. Thus each singular noun in the treebank was examined as to the proportion of its usages carrying no determiner. To accomplish this, all nouns were identified that occurred as the final word in a base NP (the final word of a base NP, if a noun, is

[7] For example, *No dogs or cats are allowed* does not mean the same as *No dogs or no cats are allowed*—or perhaps the second is simply meaningless and would never occur.

Table III. NP Coordinate Constructions of the Form "D N C N" or "D N C D N" in the Penn Treebank

	D N C N	D N C D N	Total
Both nouns singular	56 (59.6%)	38 (40.4%)	94
Both nouns plural	52 (91.2%)	5 (8.8%)	57

virtually always the head); the frequency of each noun in this set was tallied, yielding a "base-NP head frequency" for each noun. Then, for each noun, the base NPs headed by that noun were counted which contained no determiner, numerical quantifier, or possessive pronoun. The proportion of these "no-determiner usages" in relation to the base-NP head frequency was calculated, yielding a "zero-determiner ratio" (ZDR) for each noun. The reasoning was that nouns frequently used without determiners would be mass nouns—or, one might say, further towards "mass" on the count-mass continuum. The question then was whether mass nouns, defined in this way, would more likely be used in a single-determiner context as opposed to a double-determiner context.

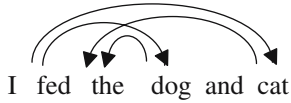
A list was made of all singular-noun tokens occurring in either SD or DD contexts, as defined above. Each word in the SD list and DD list was paired with its ZDR, to see if the word-tokens in the SD list had higher ZDRs than those in the DD list. The answer was yes: the average ZDR for SD tokens was 0.214, while the average for DD tokens was 0.117. An ANOVA (treating word-tokens as observations) showed that this difference was highly significant ($F = 6.91$, $p < 0.01$).

Before continuing, we should address one possible objection to the current argument. In a coordinate phrase of two nouns or noun phrases, the two phrases may refer to the same entity, or to two different entities. In the sentence (from the treebank) *[He] is retiring from active duty but remains a director and consultant*, it is clear that both *director* and *consultant* refer to the same entity. Let us assume that single-referent coordinate NPs usually involve singular count nouns rather than mass or plural nouns. (If this assumption is not accepted, then the objection that I am about to address fails immediately.) If so, one might argue, the preference for DD structures with singular count nouns is a kind of ambiguity-avoidance mechanism, designed to avoid the unwanted "single-referent" interpretation that would be caused by an SD structure. That is, if one wanted to refer to two different people—a director and a consultant—one might rather use the DD construction (as I just did) to avoid the single-referent implication that would arise from the SD construction. This objection seems unlikely to be valid, however, for the following reasons.

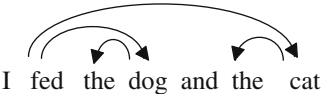
Single-referent SDs are actually quite rare; out of the 56 singular SD constructions found, only 3 were single-referent. It seems implausible to argue that people avoid using a certain structure to mean interpretation A for fear that it might imply interpretation B, given that when they do use it, they much more often intend interpretation A than interpretation B. Moreover, DD structures are not entirely incompatible with a single-referent interpretation; for example, the DD phrase *a director and a consultant*, with a single-referent interpretation, also occurs in the treebank. There is not enough data in the treebank to resolve these issues empirically, but it seems very unlikely that the avoidance of single-referent interpretations explains the correlation observed here.

The tests just reported show that SD constructions are used relatively more often with plural nouns than with singular nouns, and relatively more often with singular mass nouns than with singular count nouns. Both of these tests point to a tendency to use the SD construction more with nouns that do not require determiners. If so, what does this prove? It can be seen that, according to the multi-head theory, SD constructions (like (5)) involves crossing dependencies, whereas DD constructions (like (6)) do not:

(5)



(6)

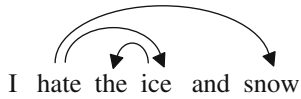


Crossing dependencies are known to be rare in English and highly constrained in their use; in light of this, we might suppose that they would cause some processing complexity.[8] By this reasoning, we might expect that people would generally avoid SD constructions, using the DD construction instead. Now, it may be noted that the correct analysis of all of the SD phrases studied here involves crossing dependencies—not just count-noun cases, but plural- and mass-noun cases as well. (Recall that cases where the determiner did not apply to the second noun were excluded from the test set.) However, with a phrase like *the ice and snow*

[8] The complexity of crossing dependencies in English is a difficult and unexplored issue. As already noted, they are generally assumed to be rare, and some grammars prohibit them entirely. But Bach, Brown, and Marslen-Wilson, (1986) showed that crossing dependencies in Dutch were less difficult to process than center-embedded structures in German, suggesting that crossing dependencies may not always be cognitively taxing.

or *the dogs and cats*, a non-crossing interpretation like (7)—where the determiner connects only to the first noun—is at least syntactically correct, even if it is not the intended interpretation.

(7)



In the case of singular count nouns, such a structure would be incorrect, since the second noun would have no determiner. Perhaps parsing involves an initial stage of finding a dependency structure that is syntactically acceptable, satisfying the "requirements" of each word (a singular count noun requires a determiner, a plural or mass noun does not).[9] Once such a structure is found, it may be modified in a later stage (perhaps in ways that involve crossing dependencies), or perhaps further interpretation does not involve dependencies at all. This explanation is admittedly somewhat complex and involves several conjectural leaps; but at present, it is the only explanation on offer for the relative avoidance of SD structures with singular count nouns.

GENERAL DISCUSSION

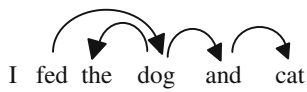
The use of coordinate phrases in written English, as reflected in the Penn Treebank, provides strong evidence for the multi-head model of dependencies in coordinate phrases. It was asked, first, whether the multi-head model predicts ordering preferences for conjuncts with regard to length, on the assumption that orderings involving greater dependency length should be less favored. The general short-long preference for right-branching structures is exactly as predicted by the multi-head model. The fact that a short-long preference is found also for left-branching structures is not as easily reconciled with the model; but syntactic choices are often guided by general strategies, and given that most dependencies in English are presumably right-branching rather than left-branching, the multi-head model predicts that a general short-long strategy would be more beneficial than a general long-short strategy. Turning to the case of determiner duplication, it was observed that single-determiner constructions are more often used with plural or mass nouns than with singular count nouns. According to the multi-head model, the single-determiner construc-

[9] Hudson (2003) has offered a similar proposal to account for certain phenomena of adjunct preposing in English.

tion involves crossing dependencies (which are presumably cognitively difficult) with singular count nouns if the words' syntactic needs are to be satisfied. Thus the multi-head model offers an explanation for why single-determiner constructions might be relatively less favored with singular count nouns.

So far, the "multi-head" model of dependency structures in coordinate phrases is the only one that has been considered. While dependency structures in coordinate phrases have not received much discussion, one other proposal deserves mention, put forth by Mel'cuk (1987). Mel'cuk proposes that the head of a coordinate phrase is the first conjunct; the conjunction is a dependent of the first conjunct, and the second conjunct is a dependent of the conjunction. Let us consider how well this "head-first" model accounts for the data presented above. With regard to conjunct ordering, the model's predictions are complex. (Here we assume, as usual, that constructions with longer dependencies are less favored.) If we assume that the conjunct-heads are at the left end of their conjuncts, then the model predicts a short-long ordering of conjuncts, for both right-branching and left-branching coordinate phrases—exactly as observed in the data. The problem is that this prediction depends on the position of the heads within the conjuncts; if, for example, the heads are at the right end of the conjuncts, then a long-short ordering is favored for left-branching coordinates. While many English constituent-types tend to have the head at the beginning, in noun phrases (which provide most of the data for left-branching coordinates) the head noun is often not at the beginning, and is often preceded by a determiner and pre-modifiers. Thus the fit of the head-first model to the treebank data is unclear (and is currently the subject of further investigation). Turning to the evidence from determiner duplication, the head-first model fares poorly. The model does not predict crossing dependencies under any circumstances; all words connect to the coordinate phrase via the first conjunct head.

(8)



Thus the model offers no explanation for the relative avoidance of single-determiner constructions with singular count nouns. Comparing the head-first model and the multi-head model overall, then, the verdict from conjunct ordering is presently unclear, while the verdict from determiner duplication clearly favors the multi-head model.

Another model that deserves consideration here is Hawkins's early immediate constituent (EIC) theory (1994). Like Gibson's DLT, the EIC theory attempts to account for a wide variety of phenomena in language

perception and production. The essential idea of Hawkins's theory is that the comprehension of a syntactic constituent is easier if its daughter constituents can be quickly identified. In effect, this results in a preference for short-long ordering of constituents; in a verb phrase containing V-NP-PP, for example, the window within which the three daughters can be identified (known as the Constituent Recognition Domain) will be smaller if the shorter of the two arguments is first. Since the EIC predicts short-long ordering in both left-branching and right-branching cases, it appears to model the data for conjunct ordering quite naturally. An interesting test case to examine in this regard is coordinates containing three conjuncts. Consider a right-branching coordinate with three conjuncts of lengths 2, 4, and 6. The DLT, combined with the multi-head theory, predicts that they will be ordered from shortest to longest (2–4–6); this is the ordering that will minimize the total dependency length. Under the EIC, however, what matters is the length of the Constituent Recognition Domain; this favors placing the longest conjunct last, but expresses no opinion about the ordering of the other two. (Whether the order is 2–4–6 or 4–2–6, the span between the beginning of the entire phrase and the beginning of the third conjunct is the same.) The treebank data was again analyzed, looking at coordinate phrases of all types containing exactly three conjuncts; 1855 cases were found. The average lengths of the three conjuncts were 3.41, 3.64, and 4.07, respectively. Since the second conjunct is longer than the first, this would seem to favor the multi-head / DLT model rather than the EIC model (though the difference fell just short of significance: $F = 3.4$, $p = 0.06$). Thus, while the data for conjunct ordering is accommodated by both theories (perhaps more easily by the EIC), the DLT has greater success in predicting ordering preferences for three-conjunct coordinate phrases.[10]

Beyond providing evidence for the multi-head theory of coordinate dependencies, this study—if its conclusions are accepted—also has implications for several other issues in linguistics and psycholinguistics. First of all, it provides further testimony—adding to the findings of Gibson

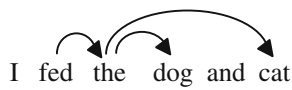
[10] Hawkins considers various cases of three-constituent constructions, and acknowledges that the basic EIC is unable to predict the shorter-to-longer ordering preference that is generally found in such cases. A special "left-to-right" procedure is proposed for handling such phenomena. Regardless of the merit of this left-to-right procedure, three-constituent ordering preferences such as these seem to favor the DLT, as it requires no extra apparatus to account for them.

A further general advantage of the DLT over the EIC—though not directly related to the data presented here—is the ability to explain the long-short preference in head-final languages; again, the basic EIC theory does not predict this, and Hawkins must account for it by quite a different mechanism.

and others—for the explanatory power of dependency structure and its connection with syntactic complexity, showing how more complex dependency structures tend to be avoided in situations of syntactic choice. It is not obvious why syntactically complex structures are avoided; broadly speaking, there are two possibilities. One possibility is that complex structures are avoided for the sake of the perceiver, on the grounds that they are difficult to comprehend (Bever, 1970; Hawkins, 1994; Temperley, 2003); another possibility, more favored in recent years, is that certain structures are in some way difficult to produce (Wasow, 1997; Arnold *et al*., 2000; Ferreira & Dell, 2000). Certain types of syntactic choice, such as those motivated by ambiguity-avoidance, seem most readily explained in perceptual terms; on the other hand, Arnold *et al*. (2000) argue that heaviness and newness effects are at least partly due to production factors. The medium may make a difference: while complexity effects in syntactic choice have been observed in both speech and writing, production limitations seem more plausible in the case of speech, since the computational restrictions are greater in that case; in writing, there is less need to produce output quickly and more opportunity to modify and revise. But this argument is not decisive, and it is certainly possible that the phenomena observed here are due at least in part to producer-based limitations.

The current study also sheds light on a basic issue in dependency grammar: the head of the noun phrase. While many have taken the main noun of a noun phrase to be its head (Bresnan, 1982; Mel'cuk, 1987; Pollard & Sag, 1987; Gibson, 1998), others have assumed that the determiner, when present, is the head of the noun phrase (Abney, 1987; Hudson, 1990; Radford, 1997). Note that under the "determiner-headed" view of the noun phrase, there are no crossing dependencies in single-determiner constructions; in (9), every word's syntactic requirements may be met, without any crossing dependencies.

(9)



Thus the determiner-headed view of noun phrases seems unable to account for the avoidance of such constructions with count nouns. This would appear to be an important empirical point in favor of the "noun-headed" model of the noun phrase.

As noted earlier, the multi-head theory of coordinate dependencies is also of interest as it seems to violate some widely-held assumptions about dependencies (at least in English): that each word has exactly one head, and that dependencies do not cross. In a single-determiner coordinate construction like that in (5), the determiner has two heads (both nouns), and

the dependencies between the nouns and the determiner and verb cross over each other. It has been suggested here that the crossing of dependencies in single-determiner constructions adds complexity, which causes such constructions to be disfavored (though not entirely avoided). It would be interesting to examine this prediction in other cases. Under the multi-head view, as noted earlier, crossing dependencies will occur any time the conjuncts of a coordinate phrase make multiple connections to outside words—for example, if conjoined verbs connect to a preceding subject and a following object, as in this sentence from the treebank:

(10)

The current view suggests that such constructions, also, should be difficult to process and hence disfavored. The problem with testing such cases by the current method is that, unlike in the "determiner-duplication" case, there is no obvious syntactic choice involved; that is, there is no easy way of rephrasing the sentence to avoid the crossing. In the sentence above, the crossing could be avoided by repeating the object phrase, perhaps with a pronoun—*El Paso owns a petroleum refinery and operates it*—but this seems awkward and unlikely. Rephrasing such a construction to avoid crossing dependencies would be more palatable if the object phrase was extremely short: Given the sentence *Fred peeled and ate it*, the non-crossing version *Fred peeled it and ate it* is not so bad. But such usages are so rare in the treebank—either single- or double-pronoun constructions—that no conclusions can be drawn about their relative frequency.[11]

As a final point, the multi-head model—and dependency structure generally—relates in an important way to what Townsend & Bever (2003) have broadly called "statistical" models of sentence comprehension. While some models have focused on syntactic or "structural" aspects of sentence processing (Frazier, 1985; Hawkins, 1994; Gibson, 1998), others have explored the role of other factors—factors of discourse structure (Crain & Steedman, 1986), frequency (e.g., the frequency of transitive versus

[11] There are also other crossing-dependency structures in English (at least under the usual assumptions about dependency structures)—for example, when the argument of a direct-object noun is separated from it by an argument of the preceding verb:

(i)

It would be interesting to know whether crossing-dependency structures such as this also cause processing difficulties.

intransitive uses of a verb—MacDonald, Pearlmutter, & Seidenberg, 1994; Jurafsky, 1996), and plausibility. Plausibility considerations might include, among other things, the thematic or semantic "fit" between a dependent and its head (Trueswell, Spivey-Knowlton, & Tanentaus 1994; MacDonald *et al.*, 1994; McRae, Tanentaus, & Garnsey, 1998). For example, Trueswell *et al.*, (1994) gave subjects sentences like those in (10), where the verb *examined* could initially be construed either as the main verb or as a reduced relative.

(11a) The defendant examined by the lawyer turned out to be unreliable.
(11b) The evidence examined by the lawyer turned out to be unreliable.

Reading times indicated that subjects were relatively more biased towards the main-verb interpretation in (11a) than in (11b). According to Trueswell *et al.*, this is explained by the fact that *defendant* is a thematically and semantically plausible subject for *examined*, while *evidence* is not. While models of this type are not generally presented in terms of heads and dependents, it should be clear that they have a close connection with the idea of dependencies. That is to say, if parsing involved creation of a dependency structure—a matching of heads with their dependents—the consideration of such word-to-word compatibilities, based on plausibility factors such as thematic or semantic fit, could occur quite naturally.[12] And here, the question of the dependency structure of coordinate phrases becomes important. Consider the sentence

(12) The defendant and the evidence examined by the lawyer turned out to be unreliable.

Under the multi-head model—and assuming that dependencies provide the word-to-word relationships on which plausibility is evaluated—either the word-pair *defendant-examined* or the word-pair *evidence-examined* could affect the overall plausibility of the main-verb reading. Under Mel'cuk's head-first model, only the first conjunct (*defendant* in this case) is connected with the verb, thus only the *defendant-examined* relationship should affect plausibility.

The extent to which such lexical dependencies are involved in ambiguity resolution is an unresolved issue. But to the extent they are, this gives further importance to the role of dependencies in human parsing, and may provide another route to the empirical investigation of dependency structures.

[12] This idea has been explored in computational linguistics, most notably by Collins (1996, 2003), whose dependency-based parsing models rely largely on statistical information about the frequency of head-dependent pairs.

REFERENCES

Abney, S. (1987). The English noun phrase in its sentential aspect. PhD dissertation, MIT, Cambridge, MA.

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.

Bach, E., Brown, C., & Marslen-Wilson, W. D. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes, 1*, 249–262.

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 279–362). New York: Wiley.

Bresnan, J. (1982). *The Mental Representation of Grammatical Relations.* Cambridge, MA: MIT Press.

Bresnan, J., Kaplan, R. M., Peters, S., & Zaenen, A. (1982). Cross-serial dependencies in Dutch. *Linguistic Inquiry, 13,* 613–635.

Brown, G., & Yule, G. (1983). *Discourse analysis.* Cambridge, UK: Cambridge University Press.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In Freedle, R. O. (Ed.), *Discourse production and comprehension* (pp. 1–40). Norwood, NJ: Ablex.

Collins, M. J. (1996). A new statistical parser based on bigram lexical dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics* (pp. 184–191). Menlo Park, CA: Association for Computational Linguistics.

Collins, M. J. (2003). Head-driven statistical models for natural language parsing. *Computational Linguistics, 29*, 589–637.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 320–358). Cambridge, UK: Cambridge University Press.

Dik, S. C. (1989). *The Theory of Functional Grammar. Part I: The Structure of the Clause*. Dordrecht: Foris Publications.

Elsness, J. (1984). *That* or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies, 65*, 519–533.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*, 296–340.

Frazier, L. (1985). Syntactic complexity. In D. R. Dowty, L. Karttunen, & A. M. Zwicky, *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives* (pp. 129–189). Cambridge, UK: Cambridge University Press.

Gaifman, H. (1965). Dependency systems and phrase-structure systems. *Information and Control, 8*, 304–337.

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.

Gibson, E., Schütze, C., & Salomon, A. (1996). The relationship between the frequency and the processing complexity of linguistic structure. *Journal of Psycholinguistic Research, 25*, 59–92.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*, 274–307.

Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge, UK: Cambridge University Press.

Hudson, R. (1990). *English Word Grammar*. Oxford, UK: Blackwell.

Hudson, R. (2003). Trouble on the left periphery. *Lingua, 113*, 607–642.

Joshi, A. K. (1990). Processing crossed and nested dependencies: An automaton perspective on the psycholinguistic results. *Language and Cognitive Processes, 5*, 1–27.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20*, 137–94.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580–602.

Levy, R. (2002). The statistical distribution of English coordinate noun phrases: Parallelism and weight effects. Paper presented at *31st NWAV Conference*, Stanford, CA, October 2002.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory of Language, 38*, 283–312.

Mel'cuk, I. A. (1987). *Dependency Syntax: Theory and Practice*. Albany, NY: State University of New York Press.

Oehrle, R. T., Bach, E., & Wheeler, D. (1988). *Categorial Grammars and Natural Language Structures*. 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.

Sleator, D., & Temperley, D. (1993). Parsing English with a link grammar. *Proceedings of the Third International Workshop on Parsing Technologies*, pp. 277–292.

Steedman, M. (1985). Dependency and coordination in Dutch and English. *Language, 61*, 523–68.

Temperley, D. (2003). Ambiguity avoidance in English relative clauses. *Language, 79*, 464–84.

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.

Townsend, D. J., & Bever, T. G. (2001). *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318.

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.

Yamashita, H., & Chang, F. (2001). "Long before short" preference in the production of a head-final language. *Cognition, 81*, B45-B55.