

Post-verbal constituent ordering in English *

Thomas Wasow
Stanford University

Jennifer Arnold
University of Rochester

1. Introduction

Although English is generally regarded as having relatively fixed word order, post-verbal constituents actually exhibit considerable variability in their ordering. (1)-(4) illustrate a number of constructions in which the ordering between two constituents is not fixed.

- (1) Heavy Noun Phrase Shift¹ (HNPS)
 - a. We take too many dubious idealizations for granted.
 - b. We take for granted too many dubious idealizations.
- (2) Dative Alternation (DA)
 - a. Kim handed a toy to the baby.
 - b. Kim handed the baby a toy.
- (3) Verb Particles (VPrt)
 - a. We figured the problem out.
 - b. We figured out the problem.
- (4) Multiple PPs
 - a. Pat talked to Chris about Sandy.
 - b. Pat talked about Sandy to Chris.

There is little, if any, semantic difference between the (a) and (b) versions in these examples; they certainly do not differ in truth conditions.

For the past few years, we have been investigating what factors influence the choice of one ordering over another in such cases, with special attention to the first two. This paper summarizes some of the factors that have been shown to matter, and discusses at somewhat greater length one factor that might be expected to matter, but appears to have little effect.

The work we report on here involves both corpus studies and psycholinguistic experiments. Much of it is described elsewhere in greater detail. Our purpose here is to pull together what is known on the topic and to try to extract some lessons from it.

2. Known factors

2.1 Weight

Behaghel (1909/10: 139) is usually credited with the first observation of what Quirk, et al. (1972) later called “the principle of end-weight”: “So bildet sich unbewußt in den Sprachen ein eigenartiges rhythmisches Gefühl, die Neigung,

vom kürzeren zum längeren Glied überzugehen [...] was ich [...] als das Gesetz der wachsenden Glieder bezeichnen möchte" [Thus, a peculiar rhythmical feel unconsciously takes shape in languages: the tendency to go from shorter to longer elements... what I... would like to designate the law of growing elements]. Behaghel (1930: 85) provides a more compact formulation: "Von zwei Gliedern von verschiedenem Umfang steht das umfangreichere nach" [Of two constituents of different size, the larger one follows the smaller one].

In a number of publications over the last few years, John Hawkins has built on Behaghel's observation, extending our understanding in a number of directions (see especially Hawkins 1994, 2000, and in press). Among Hawkins's important contributions are: documenting Behaghel's generalization with extensive corpus studies across a variety of languages; discovering that at least some head-final languages exhibit the mirror-image tendency (that is, with long constituents preceding short ones); and, most importantly, offering a detailed and plausible explanation for these ordering preferences in terms of processing constraints.

The literature contains many different proposals for precise definitions of what we are calling "weight" here. Wasow (1997) tested eight proposed definitions against corpus data, to determine what kind of characterization made the most reliable predictions about constituent ordering in English. He concluded that:

- Graded characterizations account for more of the data than categorical definitions.
- It is the relative weight of the postverbal constituents that influences their ordering, not the weight of any one constituent.
- Several proposed measures (viz. number of words, number of nodes, and number of phrasal nodes) are so highly correlated that it is impossible to choose among them on empirical grounds.

These observations provide some justification for Hawkins's practice, emulated by others (ourselves included), of using differences in the lengths of constituents as one parameter in studies of factors influencing ordering.

It is worth noting, however, that the use of length as the measure of weight was challenged decades ago. Chomsky (1975: 477) wrote:

While [...] both [*the detective brought in the suspect*] and [*the detective brought the suspect in*] are grammatical, in general the separability of the preposition is determined by the complexity of the NP object. Thus we could scarcely have [...] *the detective brought the man who was accused of having stolen the automobile in*

It is interesting to note that it is apparently not the length in words of the object that determines the naturalness of the transformation, but, rather, in some sense, its complexity. Thus "they brought all the leaders of the riot in" seems more natural than "they brought the man I saw in." The latter, though shorter, is more complex[.]

Wasow (1997), looking for a single criterion of weight, found that length worked as well as any other (and better than some), but he did not consider the possibility that length and complexity might be two distinct factors that both influence constituent ordering. We investigated this possibility in a questionnaire study and a corpus analysis.

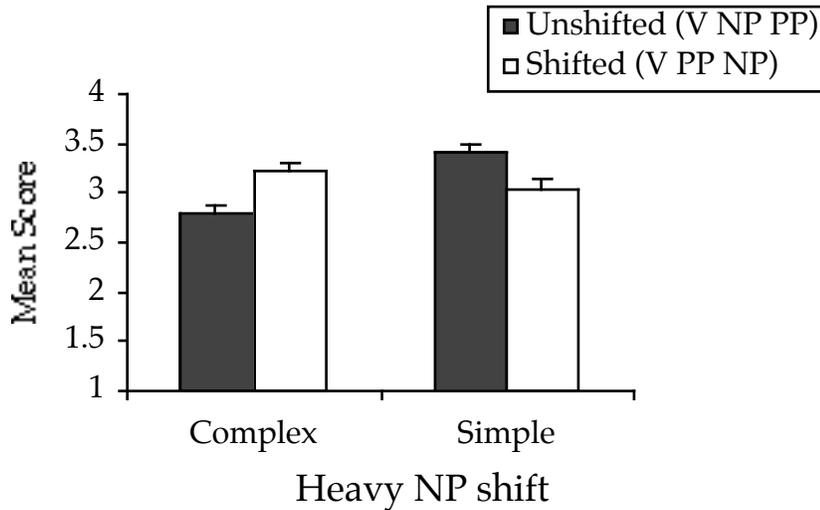
In the questionnaire study, we investigated whether complexity predicts ordering at all when length differences are controlled. We constructed sets of examples exhibiting heavy noun phrase shift, the dative alternation, and the verb-particle construction, in which the crucial noun phrases were matched by length but differed in complexity, by one natural characterization of complexity. Specifically, we constructed pairs of noun phrases that were identical in length and similar in meaning, but differed in that one contained (or consisted of) a clause, whereas the other contained none. We then put them into the constructions in question, in both the immediately postverbal position and in positions later in the clause. Hence each sentence had four variants: sentential vs. non-sentential noun phrase, crossed with immediately postverbal vs. later position. Twelve such sentence quadruples (four for each construction) were constructed, for a total of 48 test sentences. (5) gives one sample quadruple, testing heavy noun phrase shift.

- (5) a. John took only the people he knew into account.
- b. John took into account only the people he knew.
- c. John took only his own personal acquaintances into account.
- d. John took into account only his own personal acquaintances.

We used the sentences to construct four questionnaires, each containing exactly one sentence from each of the 12 quadruples, so that no subject would see more than one variant of any sentence. 88 subjects rated the acceptability of each sentence on a four-point scale: 4 for “fully acceptable”, 3 for “probably acceptable, but awkward”, 2 for “marginal, at best”, and 1 for “completely unacceptable”.

If grammatical complexity contributes to end weight, then sentences containing complex noun phrases (defined here as those containing a clause) should receive higher scores when those noun phrases occur in final position than when they occur in non-final position, and the corresponding sentences with simple noun phrases should receive lower scores when the noun phrases occur earlier than when they occur later. The following figure gives mean scores from our questionnaires for the heavy noun phrase shift examples, consolidated according to sentence type.

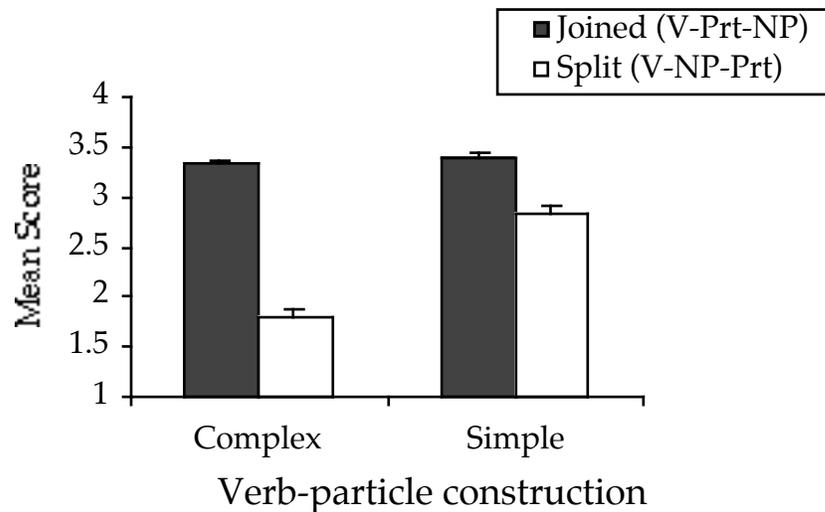
(6)



The two columns on the left show that for complex NPs, subjects rated the construction higher when it exhibited heavy noun phrase shift than when it did not. The columns on the right show that for simple noun phrases, heavy noun phrase shift reduced the average score. This is just what would be expected, if complexity does contribute to weight independently of length (and if we assume that heavy noun phrase shift is a marked construction).

The corresponding graph for the verb-particle construction is given in (7).

(7)

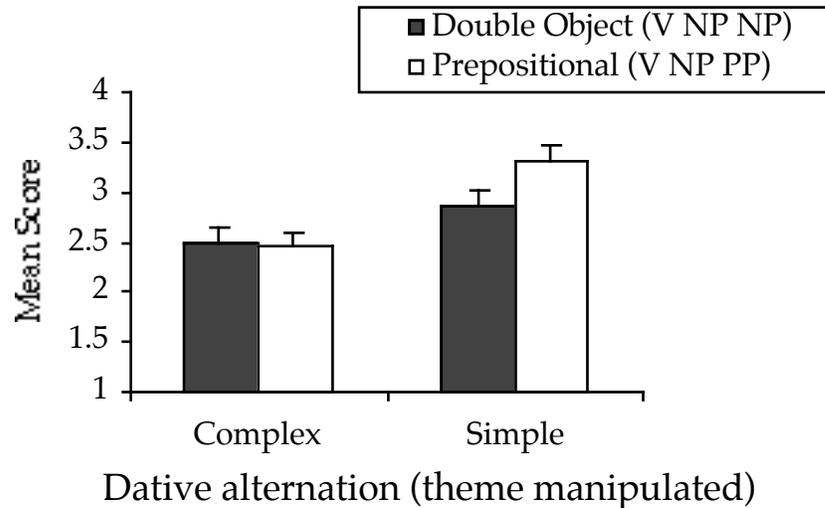


In this case, there is a reliable preference for the joined construction irrespective of the complexity². However, the difference between the two columns on the left is over twice that between the two columns on the right.

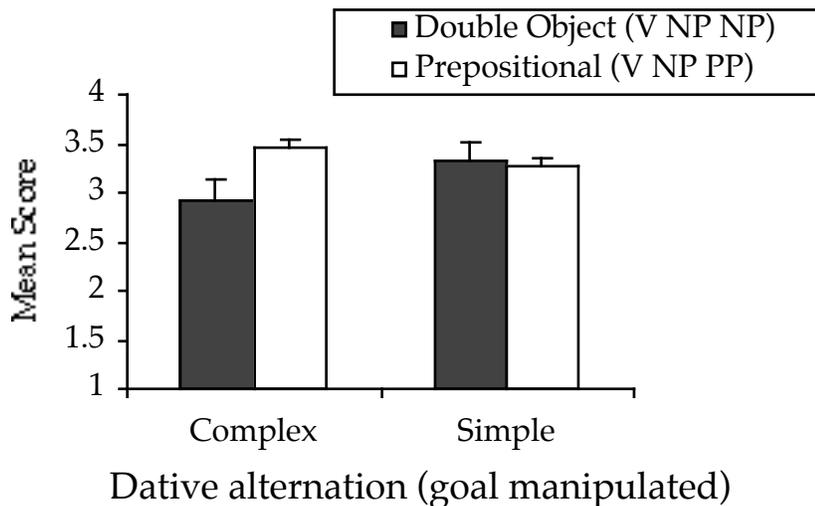
Hence, these numbers also lend support to the idea that complexity is a factor in weight independently of length.

The data regarding the dative alternation must be divided into two subsets. This is because half of the dative alternation sentences in the questionnaires varied the complexity of the theme noun phrase, keeping the goal noun phrase constant, whereas the other half varied the complexity of the goal, keeping the theme constant. The scores for these two subsets of the dative alternation data are given in (8) and (9)³

(8)



(9)



Each of these figures also suggests that complexity of the objects influences which ordering subjects will prefer. The direction of preference is what the principle of end weight predicts.

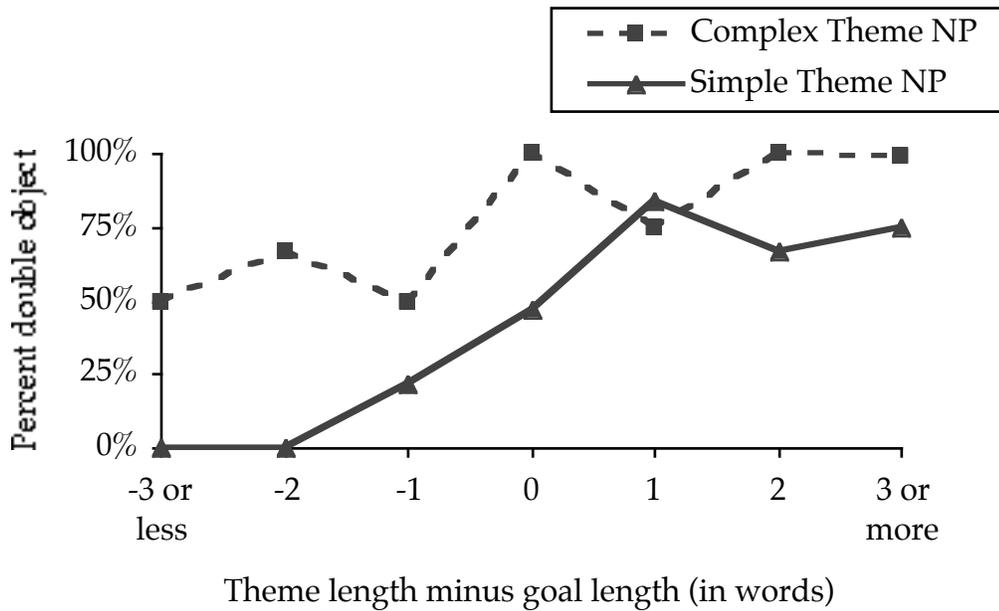
The figures above are based on numbers that are composites, averaged across multiple sentences. If the sentence quadruples are examined individually, the same pattern is revealed. In eleven of the twelve cases, the complex noun phrase shows a stronger preference for final position than the simple noun phrase. To be a bit more precise, we will use the following abbreviations: F_C represents the mean score for a sentence with a complex noun phrase in final position; N_C represents the mean score for a sentence with a complex noun phrase in non-final position; and F_S and N_S represent the corresponding numbers for a simple noun phrase. Then, for all but one of the twelve quadruples, $F_C - N_C > F_S - N_S$. Analyses of variance also reveal that complexity significantly affects preferences in ordering.⁴ These results provide support for Chomsky's original intuition that weight effects could be influenced by grammatical factors independent of length.

Thus, the questionnaire data show that when length does not predict a preference in ordering, complexity has an effect. But what happens when complexity and length both vary? We also investigated the relative roles of complexity and length in several corpus analyses.

We analyzed examples of both heavy NP shift and the dative alternation from the Aligned-Hansard corpus.⁵ For each example, we coded several properties of each constituent, including a) length in number of words, and b) whether it had post-head modification (complex) or not (simple). For both heavy NP shift and the dative alternation, variation in constituent ordering was better accounted for by looking at *both* length and complexity than by looking at either alone.

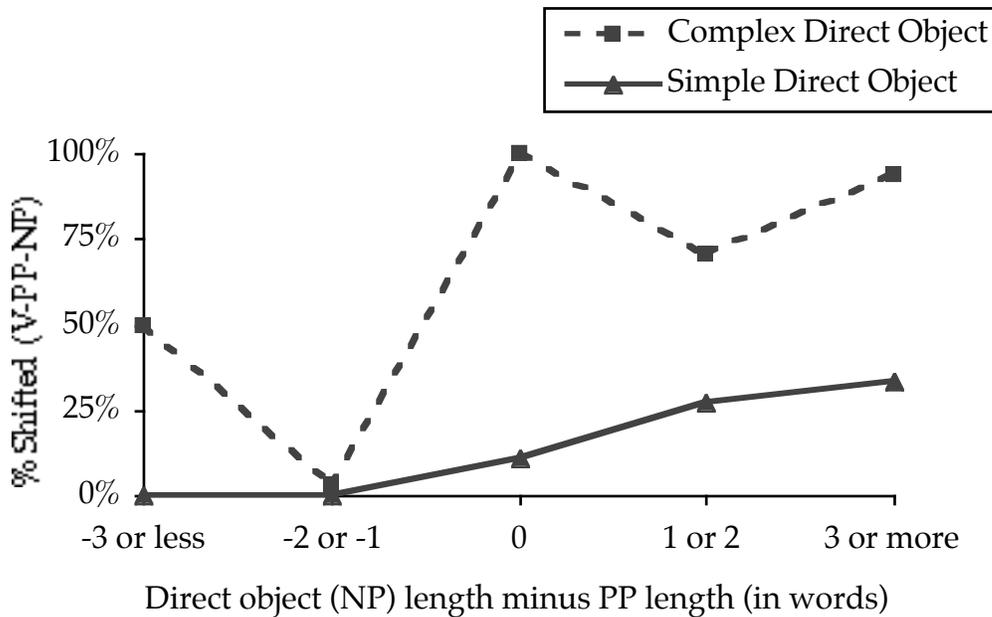
However, the contribution of each factor worked differently. For both constructions, the relative length of the two constituents predicted ordering better than the length of either one alone. But for complexity, only the complexity of the direct object (theme) noun phrase accounted for a significant portion of the variation. (10) and (11) show how constituent ordering in the dative alternation and heavy NP shift is influenced by both the relative length and the complexity of the direct object noun phrase.

(10)



Length vs. complexity: dative alternation

(11)



Length vs. complexity: heavy NP shift

We tested the reliability of these patterns for each data set with a logistic regression. A logistic regression checks each of a number of factors for whether

they significantly influence an outcome (constituent ordering, in this case), given that other correlated factors may also have an influence. That is, this logistic regression tested whether two correlated factors, length and complexity, each contribute to ordering, or whether one looks predictive only because it is correlated with the other factor.

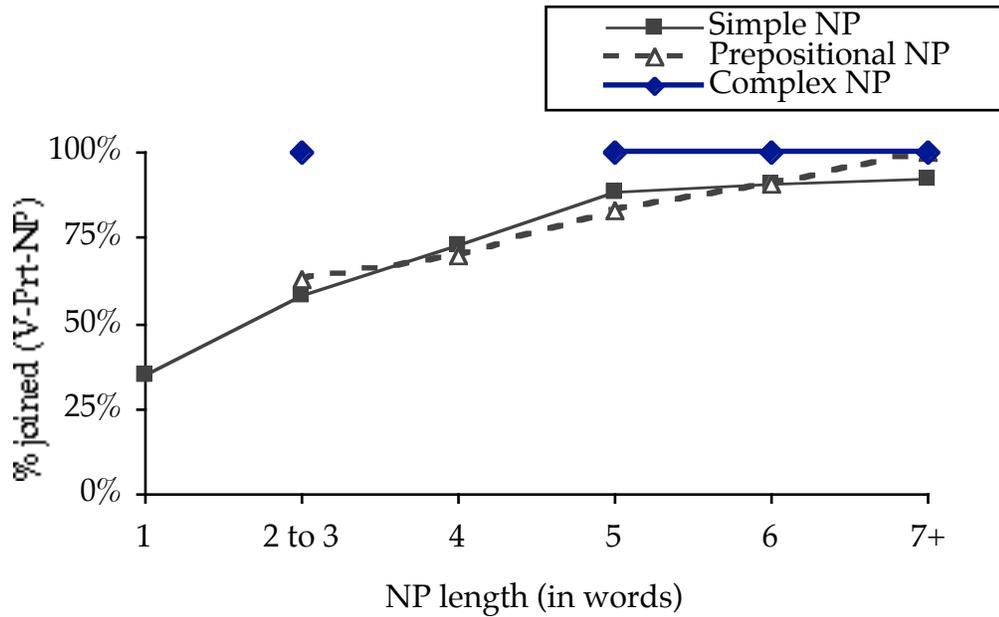
For both heavy NP shift and the dative alternation, the direct object (theme) occurred adjacent to the verb more often when it had fewer words in it than the prepositional phrase (goal) and when it was simple rather than complex ($p < .001$ for both results). The complexity of the prepositional phrase (goal) did not significantly predict ordering (p 's $> .1$).⁶

We also conducted two corpus analyses to test the effects of length and complexity on ordering in verb-particle constructions. The first investigated written corpora, the Brown and Wall Street Journal corpora, and the second investigated a spoken corpus, the Switchboard.

We analyzed 2367 verb phrases containing a particle and a direct object noun phrase from the Brown and Wall Street Journal corpora, and 1019 examples from the Switchboard corpus. We excluded items where the NP was a pronoun, since these obligatorily occur with the particle following the object. We coded each item in our sample for three properties: first, whether the particle was adjacent to the verb ("joined") or after the direct object ("split"); second, whether the object noun phrase contained a verb ("complex") or not ("simple"); and third, the length of the object noun phrase. The results were again submitted to logistic regression analyses.

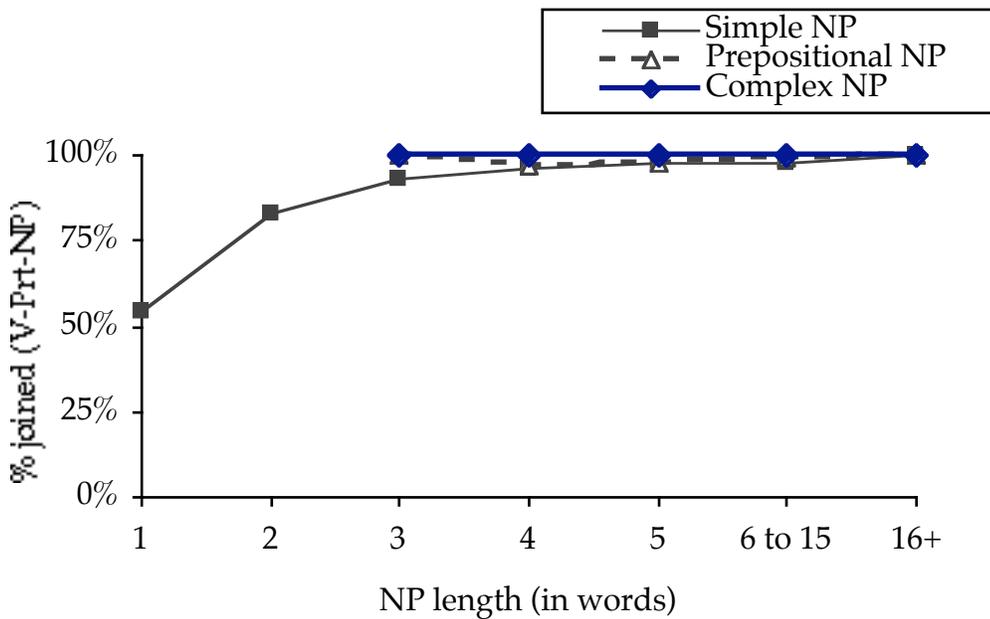
For both written and spoken corpora, the results were the same. Length significantly contributed to ordering ($p < .001$), but complexity did not. The results, presented in (12) and (13), show that every one of the items with complex noun phrases occur in the joined construction. However, almost all of these noun phrases have more than 3 words. Because the particle is such a light constituent, constructions with noun phrases longer than 3 words almost always occur in the joined construction, regardless of the noun phrase's complexity. Thus, while complexity may contribute to the heaviness of noun phrases, it is difficult to perceive its effect in the Verb-particle construction. These results match those of the questionnaire, in which the interaction of complexity and order in the Verb-particle construction was only significant in the subjects analysis.

(12)



Length vs. complexity: verb-particle construction (spoken corpus)

(13)



Length vs. complexity: verb-particle construction (written corpora)

The results from the questionnaire and corpus analyses suggest that constituent ordering is sensitive both to simple length and to a more structure-dependent criterion of complexity, at least for heavy NP shift and the dative alternation. This is not surprising if, as many people have argued, the explanation for weight effects lies at least in part in the facilitation of production and parsing. Planning or comprehending long expressions requires more memory than planning or comprehending short ones. But more highly structured expressions likewise require more resources to produce and parse. Although length and structural complexity are highly correlated, when they are teased apart, it becomes apparent that in some cases both influence constituent ordering.

2.2 Information structure

One other factor influencing constituent ordering has been discussed in the literature for many decades, namely what we will call “information structure”. A variety of terminologies have been proposed to characterize the relevant informational distinctions. One such distinction is between given, or old information, and new (e.g., Chafe, 1976; Gundel, 1988; Prince, 1992). This generalization was also succinctly stated by Behaghel (1932: 4): “es stehen die alten Begriffe vor den neuen” [old concepts come before new ones]. Gundel (1988; 229) labeled this generalization the “given before new principle” and formulated it as: “State what is given before what is new in relation to it.”

It is surprising that, despite the voluminous literatures on both weight and information structure, those two literatures are nearly disjoint. References to information that has already been introduced into the discourse can be short, since it is already familiar to the interlocutors. Pronouns, for example, whose principal function is as pointers to something already in the discourse, are single words – in fact, typically short words. Hence, one would expect that length and newness of information would tend to correlate. And this raises the question of whether weight and information structure are two distinct factors influencing ordering, or whether one of them only looks like a causal factor because of its high correlation with the other one.

Arnold, et al. (2000) investigated this question using both psycholinguistic experimentation and a corpus study. The results of the two methods agreed: neither the length of noun phrases nor the discourse status of their referents could account for constituent ordering as well as the two combined. The experiment also showed that constituent ordering was correlated with disfluencies, which are an indication that the speaker is having trouble with some aspect of production. These results together suggested that constituent ordering is influenced by constraints on planning and production: speakers tend to begin their utterances with constituents that are easier to produce, and save the more difficult constituents for later in the utterance. This strategy would also yield the short-before-long and given-before-new patterns: short NPs with given referents tend to be easier to produce than longer NPs with new referents (for a similar proposal see Bock and Irwin, 1980).

This research clearly showed that information status affects constituent ordering. It further suggested that the given/new distinction is relevant because it is one categorization of discourse status, which is one determinant of the accessibility of discourse entities. When an entity has been mentioned, it is present in the speaker's mental models of the discourse, so it is more accessible than entities that have not been mentioned. There are still, however, related questions that need further investigation. For example, there are other cues that can affect accessibility as well. While given contrasts with new, there are several types of given information, and items that have been mentioned recently or in prominent positions are perceived as more accessible than other entities (see Arnold [1998] for a review). These finer-grained distinctions in discourse status are also likely to affect ordering preferences, but to our knowledge this hypothesis has not been tested. Non-linguistic information can also affect ordering preferences. For example, speakers choose between active and passive constructions in order to begin their utterances with referents that have been made visually salient (Tomlin [1998]).

A second question for future research is whether ordering preferences are also impacted by a different dimension of information structure. Several theories of information structure use distinctions like "topic" vs. "comment", "focus" vs. "presupposition", or "theme" vs. "rheme" (see Vallduví [1992; 28-43] for a helpful survey). These terms are not identical to each other, or to the given/new distinction, but they are roughly correlated: topics/presuppositions/themes tend to be old information, and comments/foci/rhemes tend to be new information. Furthermore, what is common among these distinctions is the claim that topics/themes/presuppositions/old information tend to precede comments/rhemes/foci/new information. It is possible, however, that these ordering preferences stem from two underlying mechanisms: 1) the desire to produce given, accessible information earlier in the utterance than new, inaccessible information, and 2) the desire to put focused, important information at the end -- i.e., save the punch line for the end.

A third open question concerns cross-linguistic variation in the roles of structural and informational factors in ordering (see Siewierska [1993] for some preliminary discussion). Such unresolved issues notwithstanding, it is safe to say that the principles of end-weight and given-before-new are both operative in determining the order of post-verbal constituents in English.

2.3 Semantic connectedness

Weight and information structure do not suffice to account for constituent ordering. This is evident from the following example, from Klavans (1997: 694):

- (14) On this side of the Atlantic, the Lancaster-Oslo/Bergen corpus was designed to replicate as closely as possible the Brown corpus, the only difference being that this corpus contains British rather than American English texts.

The verb phrase *replicate as closely as possible the Brown corpus* contains three constituents: the verb *replicate*, the modifier *as closely as possible*, and the object noun phrase, *the Brown corpus*. The last two of these occur in the marked order that we are calling heavy NP shift, despite the fact that *as closely as possible* is longer and syntactically more complex than *the Brown corpus*. Moreover, this sentence occurs immediately after a full page describing the Brown corpus, so the ordering cannot be accounted for in terms of information status. Other examples in which something else must be at work are not hard to find.

What may motivate the non-canonical order in (14) is the intimate semantic connection between *replicate* and *as closely as possible*⁷. The latter is a rather generic intensifier that could modify many different expressions; and its interpretation is highly context-sensitive, depending on what it modifies. Putting this expression adjacent to the word it modifies makes the word order reflect the semantic dependency in a direct way. Once again, it was Behaghel (1932: 4) who first noted this tendency, observing that: “das geistig eng Zusammengehörige auch eng zusammengestellt wird” [what belongs together semantically is also placed close together].

Corpus evidence for the influence of semantic connectedness on heavy NP shift was presented by Wasow (1997). Over 800 examples, each involving one of five verb-preposition pairs, were classified along the dimensions of idiomaticity and constituent ordering. More specifically, on one dimension combinations of verb plus prepositional phrase were coded as non-collocations (e.g., *share that cost with others*), semantically transparent collocations (*bring the debate to an end*), or semantically opaque collocations – that is, idioms (*take our concerns into account*). On another dimension they were coded as non-shifted (*obtain prior consent from that province*) or shifted (*attribute to me a statement I did not make*). Among non-idioms (that is, non-collocations and transparent collocations), about 26% of the examples were in the non-canonical shifted order. Among idioms, the rate was about 60%. Even if the comparison is restricted to idioms vs. transparent collocations, the difference (60% vs. 47%) was statistically significant.

The numbers cited in the previous paragraph are relevant because the interpretations of the parts of an idiom like *take into account* depend on their cooccurrence with one another (see Nunberg, et al. [1994]). That is, the meaning of the whole expression is not composable from the meanings that its three words have in other contexts. In contrast, the interpretations of transparent collocations and non-collocations can be built up from the standard meanings of the parts. Hence, the verb and the prepositional phrase in an idiom are semantically more tightly connected than in a non-idiomatic verb phrase. Wasow’s corpus study shows that semantic connectedness is correlated with constituent ordering in the way that Behaghel’s observation predicts.

Hawkins (2000: 241-246) used entailment tests to determine the semantic connectedness of verbs and post-verbal prepositional phrases. Simplifying somewhat, he tested whether removing a constituent from a sentence results in a new sentence that is entailed by the original sentence. If not, then some part of the sentence is “dependent” on the removed constituent. For example, if

dropping a post-verbal prepositional phrase from sentence S results in a sentence that is not entailed by S, Hawkins says that the verb in S is dependent on the prepositional phrase; if the entailment goes through, then the verb is independent of the prepositional phrase. Thus, because *Pat ran into Chris in the park* entails *Pat ran into Chris*, but not *Pat ran in the park*, *ran* is dependent on *into Chris* but independent of *in the park*. A similar entailment test was used to test for dependency of the verb on the prepositional phrases.

Employing these tests, Hawkins coded the verbs and prepositional phrases in several hundred English sentences for dependency. He then showed that the verb and a prepositional phrase are far more likely to be adjacent if one of them is dependent on the other than if they are independent of one another. This tendency can be overridden by weight effects, but when the examples are sorted by the relative weights of the post-verbal constituents, the effect of semantic connectedness (as operationalized by Hawkins's notion of dependency) is very clear.

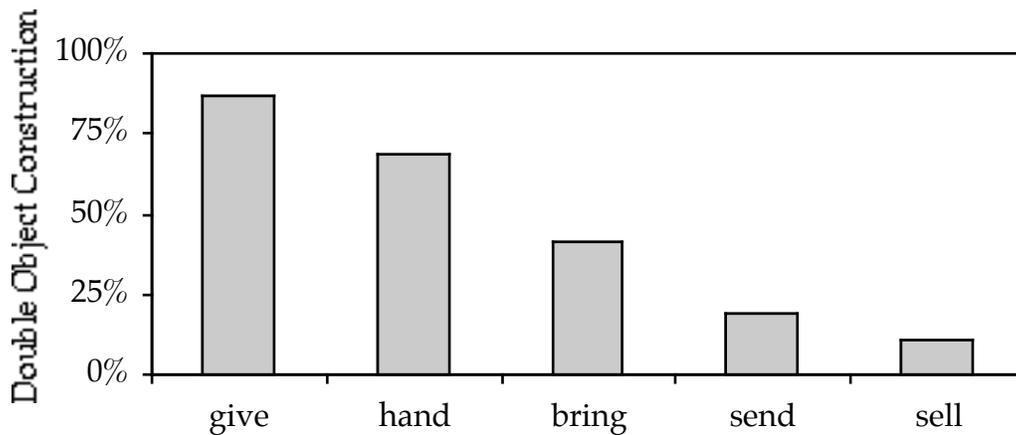
Lohse, Hawkins, and Wasow (in preparation) apply analogous entailment tests to the verb-particle construction. Preliminary results show a strong effect of semantic connectedness: dependent particles occur adjacent to the verb significantly more frequently than independent particles. Again, the weight of the object noun phrase is a stronger influence on the position of the particle, but controlling for weight brings out a robust effect of dependency.

Summing up, evidence from several quarters suggests that ordering is influenced by semantic connectedness. In particular, elements whose interpretation is dependent on the verb are more likely to occur adjacent to it than are elements that are not dependent in this way.

2.4 Lexical bias

Verbs allowing variation in ordering among their following constituents often exhibit biases towards one ordering rather than another. For example, among verbs that permit the dative alternation, some verbs occur much more frequently in the double object construction, whereas others occur more in the prepositional construction. This is illustrated in (15), which is based on samples of about 100 examples of each verb taken from the *New York Times*. (We counted only examples exhibiting one of the two constructions in question, and did not include sentences in which the theme argument was pronominal, since these require the prepositional construction).⁸

(15)



Lexical bias in the dative alternation

It is possible that there is a semantic basis for such biases. There are subtle semantic differences associated with the two constructions involved in the dative alternation (see Goldberg [1995, Chapter 2]), so perhaps the interaction of verb meaning and constructional meaning is responsible for differences like those in (15). However, these differences are subtle and seem unlikely to account for all the lexical variation in ordering preferences. Hence we provisionally conclude that speakers' mental lexicons include information about the tendency of particular verbs to appear more or less frequently in various syntactic contexts.

Several other studies also suggest that a variety of processing decisions are influenced by the frequency with which a verb appears in different contexts. MacDonald, et al. (1994) surveyed a wide range of papers dealing with syntactic ambiguity resolution and found that the apparently contradictory results could be reconciled by factoring in the lexical biases of the verbs in the stimulus sentences. In discussing the main verb vs. past participle ambiguity involved in the notorious garden path example, *The horse raced past the barn fell*, they conclude (p. 692) that "two of the types of lexical information that influence the interpretation of the [...] ambiguity are the frequencies of the past participle form and the transitive argument structure," (see also Trueswell [1996]). They arrive at analogous conclusions with respect to the other ambiguities they examined.

With respect to constituent ordering, Wasow (1997) used corpus studies to show that both the dative alternation and heavy NP shift manifest lexical biases, and argued that a verb's bias towards one ordering or another in these alternations is correlated with what other subcategorizations it permits. Stallings, et al. (1998) investigated heavy NP shift using both corpora and production experiments, likewise concluding that lexical biases exist and

correlate with alternative subcategorizations (though differing with Wasow on the details of the correlation).

In sum, individual verbs evidently differ with respect to the frequency with which the constituents following them appear in one order rather than another. These differences have been shown to influence both comprehension and production. Hence, in the absence of a demonstration that they can be derived from other factors, we tentatively conclude that lexical bias information is a distinct factor in determining constituent ordering.

3 Ambiguity avoidance

3.1 Global ambiguities

In some cases, choosing one order of constituents can avoid a structural ambiguity that would arise if another order were chosen. For example, the phrase *with a telescope* in (16a) can modify *saw*, *man*, or *hat*; but in (16b), it unambiguously modifies *saw*.

- (16) a. Pat saw a man in a funny hat with a telescope.
b. Pat saw with a telescope a man in a funny hat.

On the assumption that such ambiguities make processing more difficult (an assumption generally accepted in the sentence processing literature), speakers and writers should avoid them. One might predict, then, that unambiguous orderings would be more common than ambiguous ones (all else being equal). To our knowledge, this prediction has not previously been tested.

Using a parsed version of the Brown corpus, we extracted all verb phrases whose head verb had both a noun phrase and a prepositional phrase as sisters. After weeding out cases in which the noun phrase served an adverbial function (e.g., *arrested in Phoenix, Arizona, last week*), we were left with about 700 examples in which the noun phrase followed the verb phrases (that is, cases of heavy NP shift), and over 10,000 examples in which it did not. We then randomly selected from among the latter examples for a set of controls to compare with the shifted cases for potential attachment ambiguities.⁹

In order to make this comparison, the shifted examples needed to be coded for whether they would have been ambiguous with the other ordering – that is, with the noun phrase preceding the prepositional phrase. The non-shifted examples were coded in the form in which they actually appeared. Each example was coded for whether more than one syntactic analysis was possible and, if so, whether different analyses expressed distinct sensible meanings.

In most circumstances, a postverbal noun phrase immediately followed by a prepositional phrase allows two parses, one in which the prepositional phrase is part of the noun phrase and one in which it is not. However, when the noun phrase is a personal pronoun or a proper name, the analysis in which it also contains the prepositional phrase is impossible, or at least highly disfavored. Likewise, if the verb requires both a noun phrase object and a prepositional

phrase complement or if the prepositional phrase is part of a fixed expression restricted to the preceding verb, then only one analysis is possible. Thus, we regarded verb phrases like (17) as unambiguous.

- (17) a. [...] absent himself from his native region.
- b. [...] boosted Lincoln into the White House.
- c. [...] put pressure on the union.
- d. [...] taking the population of Nevada into consideration.

Where two syntactic analyses are possible but one of them makes no sense, and where the meanings (or at least the truth conditions) associated with the two parses are the same, we coded the verb phrases as only syntactically ambiguous. Examples are given in (18).

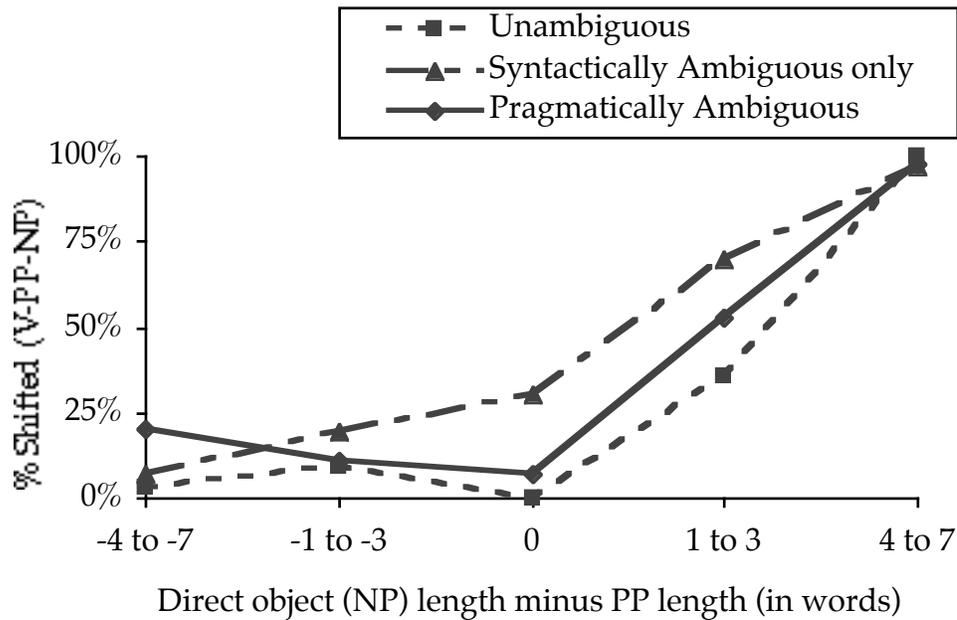
- (18) a. [...] assign a full-time maid to keeping an eye on the boy.
- b. [...] building homes on exposed coasts.

The type of example we coded as fully ambiguous is illustrated in (19).

- (19) a. [...] brought news from Kansas City.
- b. [...] censors literature only for the young.

If potential ambiguity were a relevant factor in the speaker's choice of ordering, we would expect to see more shifting in the pragmatically ambiguous items than the unambiguous items, and, possibly, more shifting in the syntactically ambiguous than unambiguous items. However, the results of this study (shown in [20])¹⁰, indicate that while length affects the likelihood of shifting, ambiguity does not.¹¹

(20)



Global ambiguity and heavy NP shift

3.2 Local ambiguities

3.2.1 Questionnaire and corpus studies

In addition to global ambiguities like the one in (16a), there are local ambiguities that can be influenced by constituent ordering. Consider sentences like (21):

- (21) a. The foundation gave Grant's letters to Lincoln to a museum in Philadelphia.
b. The foundation gave a museum in Philadelphia Grant's letters to Lincoln.

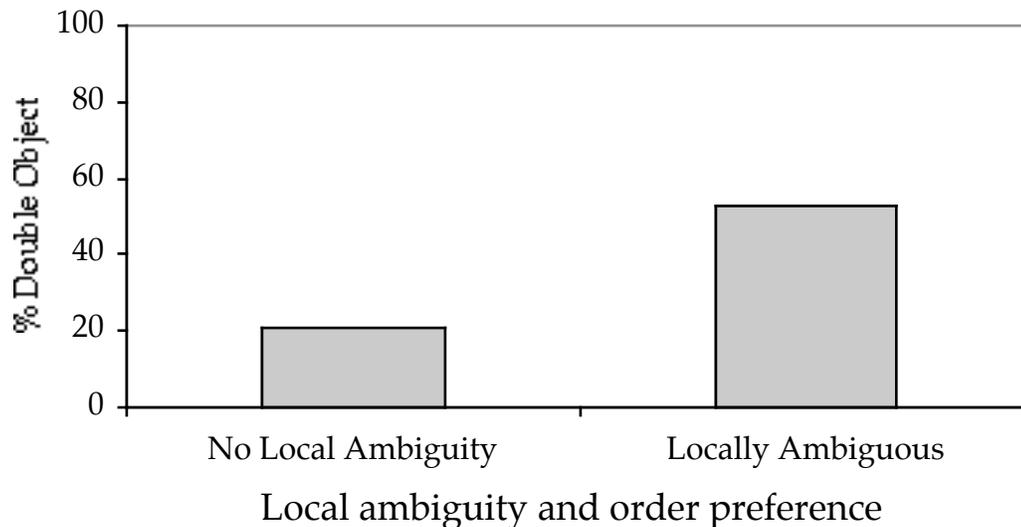
In (21a), there is a tendency for readers or listeners to misparse *to Lincoln*, analyzing it as the goal argument of the verb *gave* until *to a museum* is encountered (Boland and Bohem-Jernigan, 1998). No such misanalysis is possible in the double-object construction illustrated in (21b).

To back up our intuitions about such examples, we conducted a questionnaire study. We constructed pairs of dative alternation sentences and required participants to say which sentence in each pair seemed more natural to them. In half of the pairs, the prepositional version had a local attachment ambiguity like that in (21a) that was absent in the double-object version. The other half were unambiguous in both versions, as in (22).

- (22) a. The foundation gave Grant's letters about Lincoln to a museum in Philadelphia.
 b. The foundation gave a museum in Philadelphia Grant's letters about Lincoln.

As expected, significantly more of the participants preferred the double object version when the prepositional version contained a local ambiguity:

(23)



Given the dispreference for this kind of local ambiguity, it would be natural to suppose that speakers and writers would select constituent orders that avoid it. In particular, one might expect that forms like (21a) would be avoided in favor of forms like (21b).

We searched several corpora for examples like the sentences in (21). We found too few exactly matching one of the forms to do statistical analysis. Nevertheless, the anecdotal evidence is quite interesting.

In a search of a large corpus of *New York Times* copy, we found only three examples that were completely parallel to (21a):

- (24) a. The company gave the U.S. rights to the drug to the Population Council after declining to seek FDA approval itself, fearing retaliation from abortion opponents
 b. The Swedish drugmaker sold the rights to technology behind the product to the privately-held Colorado company Biostar in 1986.
 c. After matching the spectrogram with a phone number, the system sends a message to a telephone switch back to the customer's house and "asks" to put the call through .

Although three is a very small number, we were surprised to find any such examples in edited text, given the strong dispreference for such local ambiguities in our questionnaire study. We also found a few examples which, though not perfectly parallel to our questionnaire examples, could conceivably cause readers to misparse in a similar way. (We have underlined the ambiguous prepositional phrases).

- (25) a. Prime Minister Yitzhak Rabin of Israel has vowed to bring the question of returning the Golan to a referendum, but only after all issues are solved
- b. But ever since the catastrophic fire that killed some 80 members of the sect, the ashes of the Davidians' compound have become a dark and violent text for right-wing militants, in which many have read a call to a bitter hostility to the government
- c. A comic tells the one about a counselor's advice to a man with a fat wife [...].

These examples are actually globally ambiguous, because the attachment of the underlined prepositional phrases is not disambiguated by a subsequent *to*-phrase.

We found many more examples in the *New York Times* of sentences like (21b) – that is, double-object sentences whose prepositional counterparts would contain a local ambiguity. A few examples are given in (26)

- (26) a. Giuliani gave the commissioner the ceremonial key to the city [...]
- b. Munich would show the world a Germany in contrast to the Hitler-Nazi Germany
- c. Rick Ames , a career Central Intelligence Agency officer , has sold the Soviets the keys to the CIA's kingdom [...]

Interestingly, however, every one of the fifty-six examples of this sort that we found has a goal argument that is shorter than its theme argument. That is, the ordering they exhibit could be accounted for by Behaghel's generalization that short constituents precede long ones. Hence, they provide no evidence that ambiguity avoidance can influence constituent ordering.

We also searched the Switchboard Corpus and the Corpus of Spoken Professional American English. Although the *New York Times* corpus we used is larger by orders of magnitude, we thought perhaps that the phenomenon we were looking for would show up more in spontaneous speech than in edited text. Unfortunately, we found no examples completely parallel to those in (21). We did find a few sentences like (27), which superficially resembles (21b); but because *the mathematics* is not a possible goal of *give*, the prepositional version would not create the same kind of local ambiguity as (21a).

- (27) [...] it helps give students access to the mathematics.

As in the written corpus, all examples of this sort in the spoken corpora obey the short-before-long generalization.

In short, fairly extensive corpus searches failed to uncover any clear cases in which the choice of the double object construction over the prepositional alternative in the dative alternation could be attributed to the writer or speaker's desire to avoid an attachment ambiguity like that in (21a). In light of the strong preferences readers indicated in our questionnaire study, this is somewhat surprising. It raises the question of how strongly – if at all – speakers/writers are influenced in their choice of constituent order by the preferences of their audience.

3.2.2 Experiments

We also studied this issue by means of a production experiment, which we ran in two slightly different versions. Our objective was to induce participants to produce sentences using dative alternation verbs like *give* or *read* with theme arguments that contained a prepositional phrase, such as *Grant's letters to Lincoln* or *Grant's letters about Lincoln*, but leaving it up to participants to decide whether to use the prepositional or double object ordering.

Our experiments involved pairs of participants, whom we label the "speaker" and the "listener". These designations may be slightly confusing, because both participants speak and listen, but the crucial utterances – the ones we coded – were produced by the participants designated as speakers. The primary function of the listener was to make the experiment a communicative situation, rather than just a memorization task.

The experiment began with the speaker reading a sentence silently from a computer monitor. The sentence was not visible to the listener. These stimulus sentences were carefully designed to introduce the noun phrases we wanted the speaker to use, but in an ordering that would have to be changed in the speaker's utterance. For example, the stimulus used to try to elicit examples like (21) was (28).

(28) A museum in Philadelphia received Grant's letters to Lincoln from the foundation.

The sentence then disappeared from the screen, and the listener read a question from a list, in this case (29).

(29) What did the foundation do?

We designed the listeners' questions so that, to be responsive, the speaker would have to reformulate the original stimulus sentence, making a postverbal noun phrase into the subject and picking an ordering for the theme and the goal arguments. Half of our stimuli were like (28), in that if the speaker produced the prepositional variant of the target sentence, it would have a local attachment

ambiguity. The other half were like (30), where responses could employ either ordering without producing a local ambiguity.

(30) A museum in Philadelphia received Grant's letters about Lincoln from the foundation.

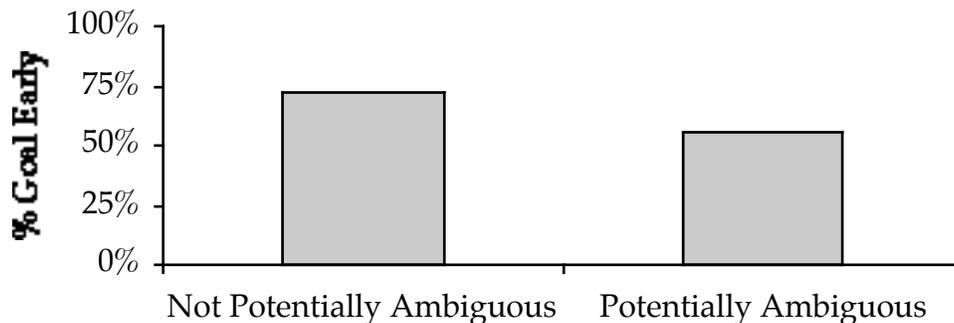
If avoidance of attachment ambiguities influences ordering, speakers should use the double-object construction more frequently in responses where the prepositional alternative would be locally ambiguous. That is, stimuli like (28) should induce a higher frequency of double-object responses than stimuli like (30). For ease of reference we will refer to stimuli like (28) as "potentially ambiguous" stimuli and those like (30) as "potentially unambiguous," although of course it is the responses (not the stimuli) that may be locally ambiguous.

Speakers were told to convey the information in the stimulus sentences as fully as possible, but we did not explicitly dictate the forms their responses should take. There was consequently considerable variation in the forms of responses; many could not be used because their form or content deviated too much from the stimuli. We coded the usable responses according to whether the goal noun phrase preceded or followed the theme. In most cases, goal-first examples were in the double object form, but we also allowed other constructions, such as those in (31).

(31) a. It gave to a museum in Philadelphia Grant's letters to Lincoln.
b. To a museum in Philadelphia, it gave Grant's letters to Lincoln.

Our initial results, summarized in (32), took us by surprise. We categorized responses in terms of whether the goal or the theme came first in the construction and whether the response was potentially ambiguous or not.¹²

(32)



Experiment I: first analysis

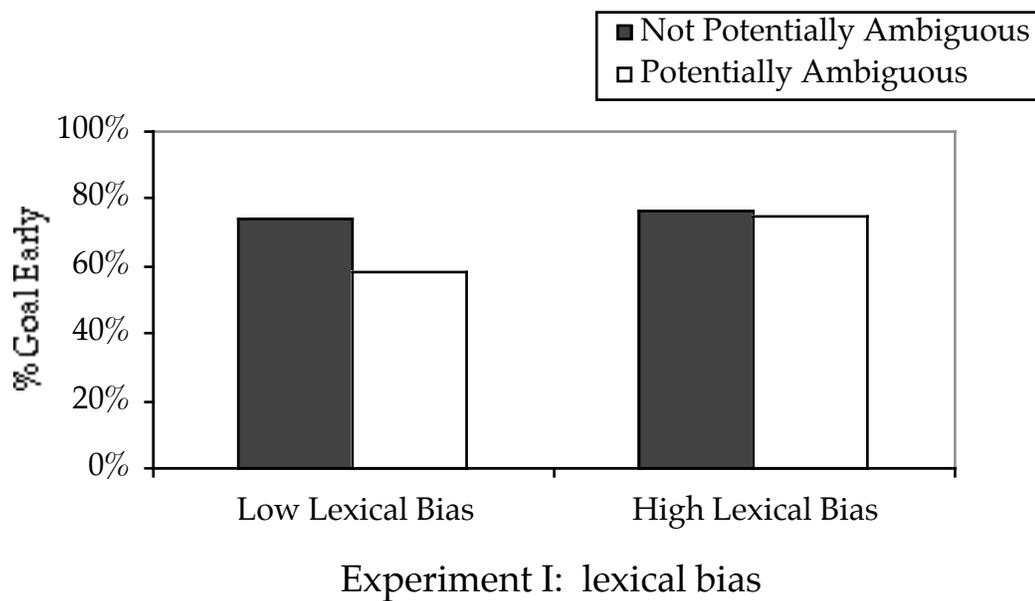
Speakers used forms with the goal early more often when their response was not potentially ambiguous -- exactly the opposite of what would be predicted if

speakers were using constituent ordering to avoid local ambiguities. However, this effect was only marginally significant in an analysis of variance of the subject means ($p=.09$).

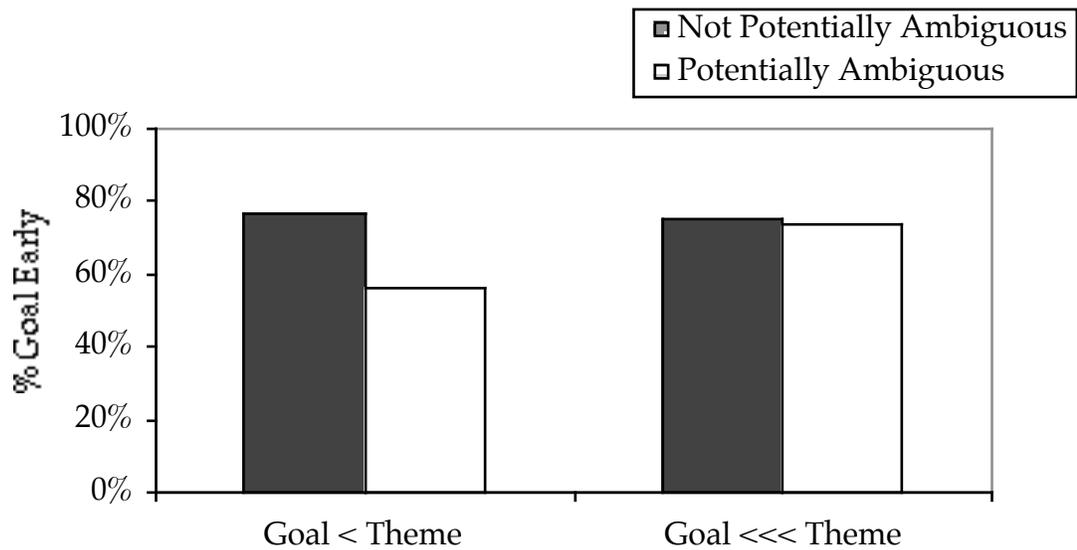
To see whether other factors might be skewing the results, we coded the responses for lexical bias of the verbs employed and for the relative lengths of the noun phrases in the responses. We based the verb biases on frequency of occurrence in each construction in a sample of the *New York Times* and classified verbs as either “high” or “low” bias towards goal-early so as to have about half of our responses in each category. Similarly, we divided our responses approximately in half, based on the difference in number of words between the theme and goal noun phrases.

The results revealed no preference for goal-early constructions in the potentially ambiguous cases. In fact, there was a slight tendency in the *opposite* direction, but this tendency was limited to cases where the lexical bias was low -- shown in (33), or the length difference between the theme and goal was small -- shown in (34).

(33)



(34)



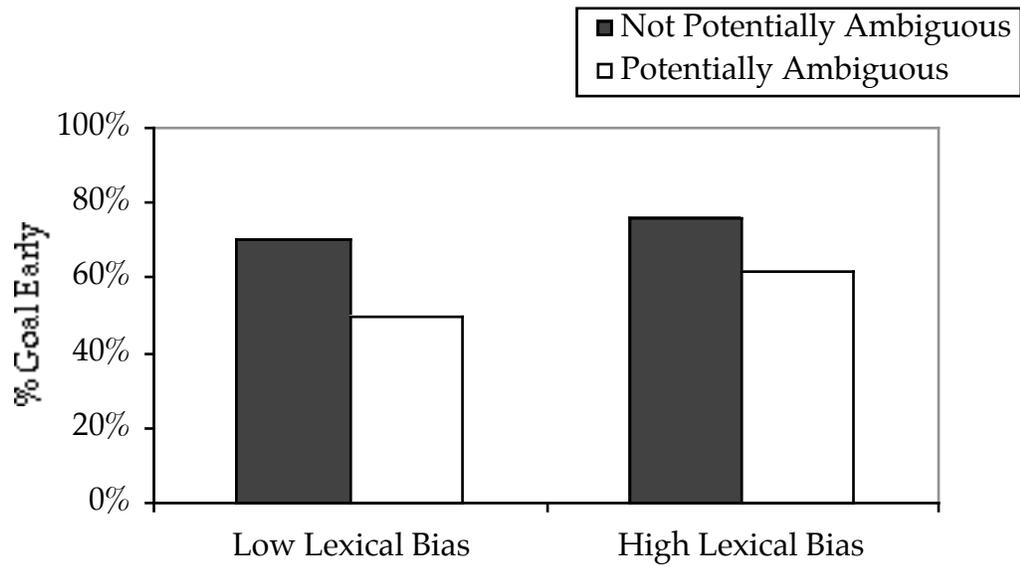
Experiment I: length effects

Notice that the theme noun phrases were generally longer than the goal noun phrases, but the magnitude of the length difference was substantially different in the two groups of responses shown in (34): an average of 2.26 words for the goal<theme group, and 4.71 for the goal <<< theme group.

We were sufficiently puzzled by the reverse ambiguity effect in our experiment that we decided to run it again, with some minor modifications. First, instead of leaving lexical bias for post-hoc analysis, we manipulated it, making sure that the lists of stimuli presented to different sets of participants were balanced for lexical bias¹³. Second, we changed the instructions to the participants to stress more emphatically that the speaker's objective should be to convey the information in the stimulus sentences. Third, we redesigned the practice session to focus the participants' attention on the need to communicate clearly. Finally we corrected a minor error in the design of the first version, which had interfered with the intended pseudo-random order of stimulus presentation.

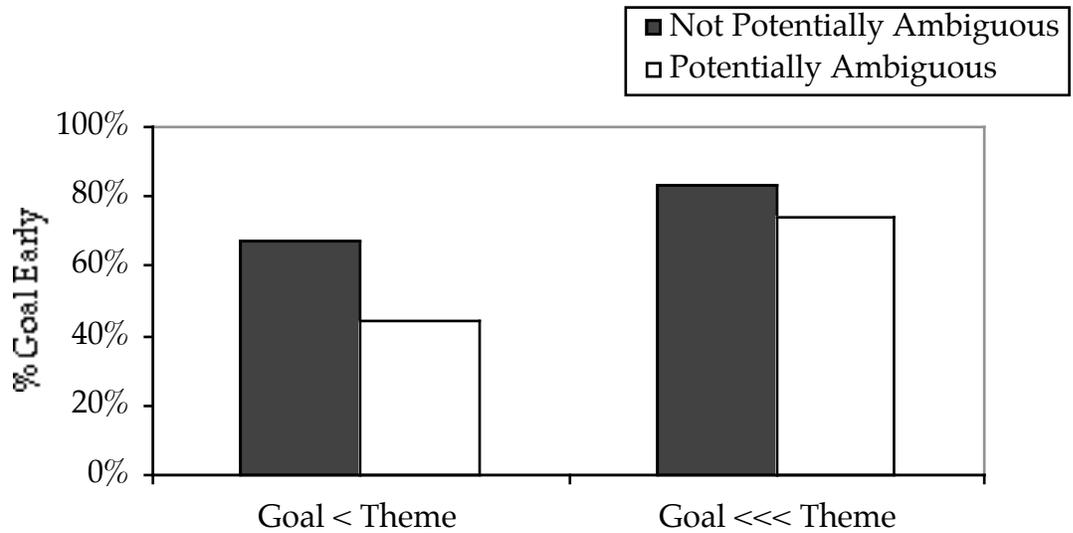
The results of the second version of the experiment are cleaner than those of the first version: the potentially ambiguous stimuli elicited significantly fewer goal-early responses across the board. While the effects of lexical bias and length are still manifest (and significant), the reverse ambiguity effect is now much clearer.

(35)



Experiment II: lexical bias

(36)



Experiment II: length effects

These experimental results strongly suggest that speakers do not use constituent ordering as a means of avoiding local attachment ambiguities. Secondly, they provide additional evidence for the role of lexical biases and

weight in ordering. Of course, our experiment only investigates one alternation in one language, so we must be cautious about drawing general conclusions. But initial indications are that ambiguity avoidance is at best a minor factor in choice of constituent order.

Our experiment naturally raises the question of why the reverse ambiguity effect occurs. We can only conjecture about the answer. One plausible suggestion¹⁴ is that a form of priming is responsible. In what we have been calling “potentially ambiguous” stimulus sentences -- that is, sentences like (28) -- the theme noun phrase (*Grant's letters to Lincoln*, in this case) contains a prepositional phrase headed by *to* and denoting a goal. This might prime the speaker to use another such prepositional phrase to denote the goal of the verb *gave*. In the “potentially unambiguous” stimuli -- sentences like (30) -- the prepositional phrase in the corresponding position plays a different role and is headed by a preposition that cannot be used in expressing the goal of *gave*. If this account of the reverse ambiguity effect is correct, then syntactic priming may be another factor influencing the ordering of constituents.

3.3 Speculations about ambiguity

Recognizing that the results of the preceding subsection require further confirmation, let us suppose for now that ambiguity avoidance is at best a minor influence on constituent ordering. Why isn't its role greater? There is a rich psycholinguistic literature suggesting that structural ambiguity makes the parser work harder (see Frazier [1985], for a particularly clear example). Hence processing -- and thus comprehension and ultimately communication -- should be facilitated if speakers avoid ambiguous structures. Alternations like those investigated in this paper often provide a ready means for speakers to avoid structural ambiguity. Furthermore, one of the best-known accounts of ordering preferences claims that constituent ordering serves to facilitate parsing (Hawkins 1994). It seems odd, then, that speakers seem to avail themselves of this method of ambiguity avoidance so little.

There are three explanations of this puzzle. First, ambiguity may not in fact greatly complicate processing, because a wealth of non-structural constraints help disambiguate utterances on the fly. Second, since other mechanisms (like prosody) may suffice for disambiguation, constituent ordering may not need to be recruited for this purpose. Third, in the cases where ambiguity does present a problem, the production system may have difficulty making use of constituent ordering to avoid the ambiguity. We believe that all three of these responses have some plausibility.

The idea that ambiguity -- especially structural ambiguity -- makes processing more difficult is so entrenched that it seems odd to question it. Explicit models of language processing consistently have to work harder on strings with more than one possible parse. There is nevertheless some reason for skepticism.

Ambiguity is pervasive in natural languages. Indeed, early computational linguists were shocked to discover that their parsers typically found many more parses in any given string than native speakers realized were possible. Because ambiguity in one region of a sentence is typically independent of ambiguity in other regions, multiple local ambiguities create combinatorial explosions that can overwhelm parsers. Assuming that languages have evolved to serve their functions reasonably efficiently, we tentatively conclude that ambiguity can't be as much of a problem for human parsers as it is for computational models.

Much of the psycholinguistic evidence suggesting that ambiguous structures add to processing difficulty is based on isolated sentences. These are commonly in written form, sometimes even lacking the usual punctuation marks. Thus, they lack the discourse pragmatic and prosodic information that would be available in connected speech. It is possible, therefore, that the stimuli in these experiments are processed quite differently from the language we encounter in everyday life.

There is a growing body of evidence (see, e.g., MacDonald, et al. [1994] and Tanenhaus, et al. [1995]) suggesting that, in extracting meaning from utterances, listeners make use of every type of relevant information as soon as it is available. Contrary to sequential modular models (e.g., Frazier and Fodor [1978], Crain and Fodor [1985]), there is no temporally separate stage of syntactic parsing; rather, phonological, syntactic, semantic, and pragmatic constraints interact to permit listeners to infer speakers' intentions. Indeed, it is doubtful whether comprehension even requires that utterances always be fully parsed.

The reason people don't notice most of the ambiguities that computational parsers reveal is that they don't make sense in the context or they don't make a difference in the context (that is, distinct parses often receive equivalent interpretations). Even isolated written sentences are spontaneously and unconsciously disambiguated if common-sense knowledge makes one interpretation highly implausible. On a recent trip, one of us observed the following signs, the first in the airport van and the latter above a moving walkway in Heathrow airport:

- (37) a. Seatbelts must be worn at all times.
b. Small children must be held securely.

Both sentences in (37) are ambiguous. (37a) could be imposing a requirement on seatbelts (that they must be worn) or on passengers (that they must wear seatbelts); likewise, (37b) could be imposing a requirement on small children (that they must be held) or on passengers (that they must hold small children). But only a linguist would notice these ambiguities, because in each case one interpretation runs counter to our common-sense knowledge of the world. Notice that the natural interpretations are not semantically parallel, but this fact presents no problem to the readers.

There is also evidence that speakers do use prosody for disambiguation (Warren, et al. [2000]), but possibly only for globally ambiguous utterances that

have two contextually plausible interpretations (Snedeker et al. [1999]). Thus, the use of diverse types of information throughout the processing of utterances makes ambiguity less of a problem for listeners than is often assumed. This being so, there is little need for speakers to employ constituent ordering as a way of avoiding ambiguities.

Even if ambiguity is occasionally a problem, it is also probable that the language production system is ill-equipped to use constituent ordering for disambiguation. This would require anticipating the full utterance and monitoring it for ambiguities **before** choosing the construction to be used. However, speakers often begin their utterances before they have finished formulating the constituents in them (Clark and Wasow [1998]), which means the construction must often be chosen before the potential ambiguity could be detected. Similar findings have been reported for other alternations which could be potentially used for disambiguation. For example, ambiguity avoidance does not play a role in the choice of whether to include optional words in utterances with potential temporary ambiguities, like *The chiropractor observed (that) you couldn't stand up straight* (Ferreira and Dell, [2000]).

4 Conclusion

Our survey has said little about why the factors we have discussed influence constituent ordering in English in the ways that they do. In some cases, plausible explanations seem fairly obvious. For example, ordering old information before new information can facilitate both production and comprehension. Old information is more accessible at a number of levels, and therefore should be easier to produce early in the utterance. This ordering also establishes the linkage to earlier sentences near the beginning of a new sentence, thereby situating the information about to be introduced in a larger context. It likewise ensures that the interlocutors share common ground that may be needed to interpret the new information. Similarly, having linear order reflect semantic connectedness is very intuitive: this sort of iconicity should facilitate both the production and the comprehension of sentences.

Other factors are not so obviously explained. Various authors (most notably Hawkins [1994]) have proposed that weight effects can be explained in terms of parsing, and Wasow (1997) argues that the principle of end-weight also facilitates utterance planning and production. Arnold et al. (2000) and Stalling et al. (1998) argue that weight affects constituent ordering by constraining planning and production. As noted above, we know of no functional explanation that has been offered for lexical biases in constituent ordering, but lexical biases have been documented in a variety of domains of production and comprehension (MacDonald et al., 1994, Stallings et al., 1998; Trueswell, 1996; Wasow, 1997).

Even in the absence of a full account of why various factors influence ordering, we believe that the exploration of what the factors are can be quite instructive. First of all, we see that constituent ordering involves the interaction of a variety of different types of factors. Such interactions are naturally modeled

in constraint-based theories of language structure and processing (see, e.g., Sag and Wasow [1999] and MacDonald, et al. [1994]). End-weight, given-before-new, semantic connectedness, and lexical biases appear to be defeasible constraints, whose influence varies with the strength of the constraint violation. The canonical ordering for a given language is likewise a set of constraints, most of which are defeasible. The preferred orderings among some specified set of constituents are those that minimize constraint violations.

Making this description into a predictive theory of ordering would require more precise formulation of the constraints and some procedure for computing the strength of each constraint. We will not attempt to provide such formulations here. Rather, we want to emphasize that even the sort of relatively informal observations provided here are relevant to controversial questions about the architecture of the human language processor.

A second lesson we draw from our examination of constituent ordering concerns the competence/performance distinction (Chomsky 1965, Chapter 1). Much work in generative grammar has been predicated on the assumption that competence – knowledge of language – consists of categorical principles, rules, or constraints. Non-categorical statistical tendencies have been assumed to reflect performance – how knowledge of language is put to use. Our studies make clear, however, that at least some categorical constraints are simply the limiting cases of more general statistical tendencies. For example, if a verb in English has a personal pronoun as its direct object, nothing can intervene between the verb and that pronoun (except, in some dialects, another pronoun), as illustrated in (38).

- (38) a. We figured it out.
b. *We figured out it.
c. They took it into account.
d. *They took into account it.
e. The president hands it to the secretary.
f. *The president hands the secretary it.

But such facts are clearly not unrelated to end weight. Pronouns are single unstressed syllables – as light as any expressions in the language. If the requirement that pronominal direct objects be immediately postverbal is part of the competence grammar of English but the principle of end weight is a performance constraint, our theory of language is missing a generalization.

Another example concerns co-occurrence restrictions. It is generally assumed that competence includes verb subcategorizations – that is, information about what kinds of complements each verb takes – but not the frequencies with which individual verbs occur in each environment. Thus, the fact that *donate* does not occur in the double object construction is assumed to be represented in a competence grammar of English, but the observation that *sell* occurs with two objects much less frequently than *give* does is taken as a fact about performance. This strikes us as arbitrary. Indeed, supposedly categorical subcategorizations are sometimes violated when there is a compelling reason to do so. For example,

the verb *begrudge* is regularly cited as “Non-Alternating Double Object Only” (Levin [1993; 47]), but (39) occurs in Eugene O’Neill’s play, “Ah Wilderness”.

(39) We don't have to begrudge it to our children.

The double object alternative to (39), **We don't have to begrudge our children it*, would violate the constraint illustrated in (38), so the usual subcategorization of *begrudge* is violated. It is by no means clear how to divide the work of explaining (39) between the competence grammar and performance factors.

More generally, we see little justification for the standard assumption that our knowledge of language, as codified in generative grammars, should be restricted to categorical information. Chomsky (1962; 128) wrote, “It seems clear that probabilistic considerations have nothing to do with grammar,” and this view has remained the dominant one among grammarians. We think it is far from clear. Indeed, the evidence suggests that people's knowledge of language does include probabilistic information.

Finally, let us close with a methodological comment. Theoretical research in syntax has relied for over forty years almost exclusively on one type of data, namely, introspective judgements of well-formedness by native speakers. This has been a remarkably productive methodology, for it has allowed syntacticians to focus on theory construction and to test hypotheses without conducting experiments in more elaborate and time-consuming paradigms. It is not without its pitfalls, however. The number of speakers consulted is usually very low (very often just one); those speakers very often know what hypothesis is being tested – and may even have a stake in the outcome; context and order of presentation of stimuli are rarely controlled; and results are assumed to be categorical, so conflicting judgements are simply chalked up to dialect differences or performance errors. Even if these methodological shortcomings were corrected (see Schütze, 1996 for discussion), it would be dangerous for an entire field of study to rely exclusively on one experimental paradigm.

It is time, therefore, for theoretical syntacticians to pay more attention to usage data. Modern technology has made it relatively easy to do this in a systematic way. We hope that this paper has illustrated how informative it can be.

References

- Arnold, Jennifer 1998 Reference form and discourse patterns. Stanford University dissertation.
- Arnold, Jennifer, Thomas Wasow, Anthony Losongco, & Ryan Ginstrom 2000 Heaviness vs. newness: the effects of complexity and information structure on constituent ordering. *Language* 76: 28-55.
- Behaghel, Otto 1909/10 Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern [Relationships between size and ordering of constituents]. *Insogermanische Forschungen* 25: 110-42.
- Behaghel, Otto 1930 Von deutscher Wortstellung [On German word order]. *Zeitschrift für Deutschkunde, Jargang 44 der Zeitschrift für deutschen Unterricht*: 81-9.
- Behaghel, Otto 1932 *Deutsche Syntax eine geschichtliche Darstellung*. Band IV. *Wortstellung. Periodenbau*. Heidelberg: Carl Winters Universitätsbuchhandlung.
- Bock, J. Kathryn and David E. Irwin 1980 Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, 19: 467-484
- Boland, Julie E. and Heather Bohem-Jernigan 1998 Lexical constraints and prepositional phrase attachment. *Journal of Memory and Language*, 39: 684-719.
- Chomsky, Noam 1962 A transformational approach to syntax. *Third Texas Conference on Problems of Linguistic Analysis of English, May 9-12, 1958. Studies in American English*. Austin, Texas: University of Texas.
- Chomsky, Noam 1975 *The Logical Structure of Linguistic Theory*. Chicago: University of Chicago Press [1955].
- Chomsky, Noam 1965 *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: MIT Press.
- Crain, Stephen and Janet D. Fodor 1985 How can grammars help parsers? In: David Dowty, Lauri Karttunen, and Arnold M. Zwicky (eds) *Natural Language Parsing*. 94-128. Cambridge: Cambridge University Press.
- Frazier, L. (1985) Syntactic complexity. In: D. Dowty, L. Karttunen, and A. M. Zwicky (eds) *Natural Language Parsing*. 129-189. Cambridge: Cambridge University Press.
- Frazier, Lyn and Janet D. Fodor 1978 The sausage machine: a new two-stage model of the parser. *Cognition* 6: 291-325.

- Goldberg, Adele 1995 *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Gundel, Jeannete K 1988 Universals of topic-comment structure. In: Michael Hammond, Edith A. Moravcsik, and Jessica R. Wirth (eds) *Studies in syntactic typology*. Amsterdam: John Benjamins
- Hawkins, John A. 1994 *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. 2000 The relative order of prepositional phrases in English: Going beyond Manner-Place-Time. *Language Variation and Change*. 11, 231-266.
- Hawkins, John A. in press Why are categories adjacent? *Journal of Linguistics* 37..
- Lohse, Barbara, John A. Hawkins, and Thomas Wasow in preparation Figuring the particle placement out.
- Klavans, Judith 1997 Computational linguistics. In: William O'Grady, Michael Dobrovolsky, and Francis Katamba (eds) *Contemporary Linguistics: An Introduction*. London: Longman.
- Levin, Beth 1993 *English Verb Classes and Alternations*. Chicago: University of Chicago Press.
- MacDonald, Maryellen C., Neal J. Pearlmutter, & Mark S. Seidenberg 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 1014: 676-703
- Nunberg, Geoffrey, Ivan A. Sag, & Thoma Wasow 1994 Idioms. *Language*, 703: 491-538
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik 1972 *A Grammar of Contemporary English*. London: Longman.
- Sag, Ivan A. and Thomas Wasow 1999 *Syntactic Theory: A Formal Introduction*. Stanford, California: CSLI Publications.
- Schütze, Carson T. 1996 *The Empirical Base of Linguistics*. Chicago: University of Chicago Press
- Siewierska, Anna 1993 Syntactic weight vs. information structure and word order variation in Polish. *Journal of Linguistics* 29: 233-265
- Snedeker, Jesse, Lila Gleitman, Michael Felberbaum, Nicole Placa, and John Trueswell 2000 Prosodic choice: effects of speaker awareness and referential context. In: *Proceedings of the Twenty-second Annual*

Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates.

Stallings, Lynne M., Maryellen C. MacDonald and Padraig G. O'Seaghdha 1998
Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language* 39: 392-417.

Tanenhaus, Michael K., Spivey-Knowlton, Kathleen M. Eberhard, and Julie E. Sedivy 1995
Integration of visual and linguistic information in spoken language comprehension. *Science*, 268: 632-634.

Tomlin, Russell S. 1998 Mapping Conceptual Representations into Linguistic Representations: The Role of Attention in Grammar. In: Jan Nuyts & Eric Pederson (eds.), *With Language in Mind*. Cambridge: Cambridge University Press.

Vallduví, Enric 1992 *The Informational Component*. New York: Garland Publishing.

Warren, Paul, Amy J. Shafer, Shari R. Speer, and S. David White 2000
Prosodic resolution of prepositional phrase ambiguity in ambiguous and unambiguous situations. *UCLA Working Papers in Phonetics*, 99: 5-33.

Wasow, Thomas 1997 Remarks on grammatical weight. *Language Variation and Change* 9: 81-105

Williams, Robert S. 1994 A statistical analysis of English double object alternation. *Issues in Applied Linguistics* 5. 37-58.

Notes

* We received assistance in the studies cited in this paper from many people, especially the Stanford students who served as research assistants. Among the people whose help we relied on are: Pete Alrenga, Cheryl Andrada, Ash Asudeh, Amy Brynolfson, Kerry Danahy, Erica Don, Matt Flannery, Ryan Ginstrom, Reiko Ginstrom, Tony Losongco, Nina Kim, Max Klee, John Niekrasz, and Jason Weinstein-Tull. We also have been greatly influenced in our thinking about these matters by our interactions with Herb Clark, Jack Hawkins, and Ellen Prince. Of course, we take full responsibility for any shortcomings.

¹ No significance should be attached to the use of this procedural terminology. This is merely the conventional name assigned to the alternation in question, and should not be taken as committing us to a transformational analysis of it.

² Corpus studies show that the verb-particle construction is highly sensitive to the length of the object noun phrase. In particular, when it is longer than a single word, the frequency of the split construction drops off dramatically. The object noun phrases in the questionnaire stimuli were of necessity at least two words long, to allow for the possibility of complexity. More precisely, in two of the sets of verb-particle stimuli, the noun phrases were two words long, and in the other two, they were three words long. The preference for the joined construction with both simple and complex objects is no doubt due to this length effect.

³ The scores should not be compared across these subsets (nor across the studies for the different constructions). Subjects were asked simply to rate sentences for acceptability. Within sentence quadruples, stimuli were matched for vocabulary, plausibility, etc. Many such factors, which are irrelevant to the issues under discussion, undoubtedly affected the scores subjects gave. Hence, comparisons across sentence quadruples are not informative.

⁴ For heavy NP shift and the dative alternation (goal manipulation), the results were significant by both subjects and items. For the verb-particle construction and Dative alternation (theme manipulation), they were significant only by subjects. The critical finding in each case was an interaction between complexity and order (heavy NP shift).

⁵ We originally collected these examples for the corpus analysis presented in Arnold, Wasow, Ginstrom, and Losongco (2000).

⁶ The Heavy NP shift data consisted of two types of constructions: *bring...to...* (e.g., *bring to the table a number of issues*), and *take...into account* (e.g., *take this issue into account*). Shifting was more frequent with *take...into account* than *bring...to...* We also tested whether the relative discourse status of the two constituents contributed to ordering, along with length and complexity. It did for Heavy NP shift ($p < .001$) but it did not for the Dative alternation ($p > .1$).

⁷ We are grateful to Jack Hawkins (p.c.) for pointing this out to us.

⁸ Arguably, these data conform to Williams's (1994: 53) claim that the double object construction is favored by verbs that require two post-verbal arguments, as opposed to verbs that can be realized as simple transitives. We are skeptical of that claim, however, as it is based on a very small sample and on the debatable assertion (p.45) that *give* requires both its theme and goal to be expressed.

⁹ We initially pulled the same number of control examples as were in our HNPS set, but some of them turned out not to exemplify the construction we were studying and had to be discarded. We therefore ended up with slightly more HNPS examples than controls.

¹⁰ (20) excludes examples in which the main verb is a form of *be*. We excluded those because they seemed to us different from the others, in that the following noun phrases are predicate nominals, not objects. We also noticed that disproportionately many of these examples have prepositional phrases that are discourse markers, such as *of course* or *in fact*. Consequently, this graph is based on about 1100 examples. A statistical analysis including the examples headed by forms of *be* is not substantially different.

¹¹ A logistic regression confirms this generalization: length (in terms of number of words) is a significant factor in the model ($p < .001$), but ambiguity is not ($p = .5$).

¹² In some cases the responses deviated in critical ways from the intended form, either introducing or removing the potential ambiguity. We report our results in terms of the potential ambiguity of the response, not the stimulus. We consider a response potentially ambiguous if it would be ambiguous in the prepositional/non-shifted form, regardless of the ordering used by the participant.

¹³ Note that the experimental design did not give us complete control over what verbs speakers used in their responses. For example, a speaker might have used *donated* or *sent* instead of *gave* to paraphrase (28). As it turned out, however, in the large majority of responses, the verb was the one we had targeted. There was actually more variation in the length of the noun phrases in speakers' responses, because it was very common for them not to repeat the noun phrases verbatim. Consequently, we were not able to manipulate the relative length of the noun phrases and had to continue with a post-hoc analysis of that variable.

¹⁴ We are grateful to Gary Dell (pc) for this idea.