HEAVINESS VS. NEWNESS: THE EFFECTS OF STRUCTURAL COMPLEXITY AND DISCOURSE STATUS ON CONSTITUENT ORDERING

JENNIFER E. ARNOLD
University of Pennsylvania

THOMAS WASOW
Stanford University

ANTHONY LOSONGCO
Yale University

RYAN GINSTROM
Ohio State University

Variations in postverbal constituent ordering have been attributed to both grammatical complexity (heaviness) and discourse status (newness), although few studies compare the two factors explicitly. Through corpus analysis and experimentation, we demonstrate that both factors simultaneously and independently influence word order in two English constructions. While past investigations of these factors have focused on their effects in language comprehension, we argue that postponing heavy and new constituents facilitates processes of planning and production.*

Even relatively fixed-word-order languages like English permit certain phrases to occur in more than one order. Examples 1–3 illustrate three familiar alternations in the sequencing of postverbal constituents.

(1) Heavy NP Shift[1] (HNPS)
  a. The waiter brought the wine we had ordered to the table.
  b. The waiter brought to the table the wine we had ordered.
(2) Dative Alternation (DA)
  a. Chris gave a bowl of Mom's traditional cranberry sauce to Terry.
  b. Chris gave Terry a bowl of Mom's traditional cranberry sauce.
(3) Verb-Particle
  a. Sandy picked the freshly baked apple pie up.
  b. Sandy picked up the freshly baked apple pie.

Why does such variation occur? This question can be asked from two closely related perspectives: (1) What factor(s) can influence the choice of one ordering over the other, and (2) What function(s) can constituent ordering variation serve?

This article addresses both of these questions. For the first question, we focus on proposals that grammatical complexity (heaviness) and discourse status (newness) are crucial factors in determining constituent ordering. Although both factors have been considered by different scholars, most studies overlook the possible effects of one or the other, and almost no studies have simultaneously investigated the two together in relation to constituent ordering. Some authors have even explicitly proposed that only one factor or the other is responsible for variation in constituent ordering. This led us to ask whether both newness and heaviness are relevant factors, or whether just one can account for constituent ordering variation. We investigated this question through

* We would like to thank Erica Don and Reiko Ginstrom for their help with the experiment. Herb Clark provided many useful discussions of the issues in this paper, as well as detailed comments on an earlier draft. Thanks also to Dennis Preston, Ellen Prince, and three anonymous *Language* referees for their comments and suggestions. Special thanks to Amy Brynolfson, coauthor on a shorter version that was presented at the 1997 annual meeting of the Linguistic Society of America, in Chicago, under the title 'Save the Worst for Last'. Thanks also to the statistical consulting service at Stanford University. The research reported here was partially supported by NSF Grant #SBR-9309612 to Stanford University; Anthony Losongco was funded through the Minority Summer Research Exchange Program.

[1] The terminology used here is for convenience and should not be taken as a commitment to analyses of these phenomena involving movement transformations.

both a corpus analysis and an elicitation experiment. Our results reveal effects of both newness and heaviness, and show that neither can be ignored.

We also discuss several possible functions of constituent ordering. Past accounts of constituent ordering have focused largely on the benefits it brings to the listener. In contrast, we consider the possible advantages that choices in constituent ordering may offer to the speaker. Our experimental data provide new evidence about the influence of production difficulty on constituent ordering.

1. Discourse status vs. structural complexity. Past investigations of variable constituent ordering have generally attempted to explain the variation either in terms of heaviness, with heavy elements coming late, or in terms of newness, with new information coming late. With few exceptions, past work has focused on only one factor at a time; here is a brief summary of the generalizations that have been offered.

1.1. Heaviness. The observation that long, complex phrases tend to occur near the ends of their clauses goes back to Behaghel (1909/10), who dubbed it *Das Gesetz der Wachsenden Glieder* (the law of the growing elements). Behaghel (1930) gives the following succinct statement of it:

> Von zwei Gliedern von verschiedenem Umfang steht das umfangreichere nach.
> ‘Of two constituents of different size, the larger one follows the smaller one.’

Essentially the same generalization is offered by Quirk et al. (1972), who label it ‘end-weight’.

Many definitions of grammatical weight have been offered in the literature. Some follow Behaghel’s original suggestion that heaviness is simply a matter of string length. Others, following Chomsky (1975: 477) have posited more sophisticated criteria of grammatical complexity, usually based on tree structures. Wasow (1997a) compares a variety of proposed definitions and tests them against corpus data. These comparisons reveal high correlations among the various characterizations and show that all are quite good predictors of constituent ordering. Put another way, the phenomenon of end weight is sufficiently robust to be manifest using almost any plausible definition of heaviness.

The studies described in this article measure heaviness as the difference in length between two constituents, in terms of number of words. This operationalization of heaviness has two advantages. First, it represents the relative weight of the two constituents. Although some discussions of HNPS focused only on the weight of the direct object NP (e.g. Postal 1974, Erdmann 1988), recent work has shown that it is the relative weight that is important (Hawkins 1994, Wasow 1997a, Stallings 1998). Second, using relative number of words permits us to observe the graded nature of weight effects (see Wasow 1997a). In all studies, our results consistently confirm that the heavier constituent tends to come later than the lighter one.

1.2. Newness. Many researchers have recognized that word order and choice of construction are correlated with information status in many languages (e.g. Birner & Ward 1998, Chafe 1976, 1994, Firbas 1966, Gundel 1988, Gundel et al. 1993, Prince 1992, Vallduvi 1990). There is a vast literature of interrelated notions about information structure and how it relates to the way that discourse information is represented in the minds of speakers and hearers, and a variety of terms have been proposed to account for informational distinctions. Some distinctions focus on the informational partitioning of propositions, for example ‘topic’ vs. ‘comment’ (Hockett 1958), ‘focus’ vs. ‘presupposition’ (Chomsky 1970), ‘theme’ vs. ‘rheme’ (Firbas 1966), or ‘focus’ vs. ‘ground’ (Vallduvi 1990). Other scholars have concerned themselves with differentiating ‘given’ or ‘old’ information from ‘new’ information, focusing instead on the information status of discourse entities, or the accessibility of information to discourse participants (Behaghel 1930, Clark & Clark 1977, Clark & Haviland 1977, Halliday 1967, Gundel 1988, Gundel et al. 1993, Prince 1981, 1992). Other scholars propose unified theories of information structure that include both informational and accessibility distinctions (e.g. Lambrecht 1994).

While there are differences in the extensions and applications of these distinctions, most investigators have argued that topics/themes/presuppositions/old information tend to precede comments/rhemes/foci/new information. This generalization has been succinctly summarized by Clark and Clark (1977:548): ‘Given information should appear before new information’. Following Gundel (1988:229), we will call this the GIVEN BEFORE NEW PRINCIPLE, or just given-before-new.[2]

It is beyond the scope of this article to evaluate the above characterizations of information status. Instead, we will adopt the distinction between discourse-given, inferable, and discourse-new, originally proposed by Prince (1992). Prince’s characterization of information structure has two advantages. First, it offers a straightforward coding scheme, which is useful for empirical studies. An NP is classified as given if its referent has been previously mentioned in the discourse. An NP whose referent has not been explicitly mentioned but could be inferred from something else that was mentioned is classified as inferable. Only NPs whose referents are truly new to the discourse are classified as new. Second, the given/new categorization of information structure has often been used to describe differences in the accessibility of discourse entities (e.g. Chafe 1994, Clark & Haviland 1977, Gundel et al. 1993). This will be relevant in considering how constituent ordering impacts the production and processing of information.

It is important to note, however, that the given/new contrast is a simplified representation of accessibility. Although new information is always less accessible than given information, there are many different types of given information; these fall at different points along a continuum of accessibility. More comprehensive accounts of accessibility have been proposed, in which phenomena such as recency of mention are considered (e.g. Ariel 1990, Arnold 1998, Birner 1998, Birner & Ward 1998, Gundel et al. 1993). This means that the accessibility of ‘‘given’’ information can differ greatly, depending on the coding scheme used by a particular study. In the two studies reported here, the samples of given constituents differ in terms of the average accessibility of their referents. The implications for this will be discussed in the conclusions.

For the purposes of this article, the term NEWNESS refers to the distinctions among given, inferable, and new information.[3] While this contrasts with the traditional term

[2] Gundel’s exact formulation: ‘State what is given before what is new in relation to it’. Gundel also proposes a second information-based generalization about ordering, which she calls the FIRST THINGS FIRST PRINCIPLE: ‘Provide the most important information first’.

In a similar vein, Givón (1988:252) proposes that speakers across languages generally ‘Attend first to the more urgent task’. But Givón also has a version of given-before-new: he suggests that constituents usually appear in order of decreasing ‘topicality’. Topicality is partly measured in terms of recency of mention, which could be considered a measure of givenness.

It should be evident that these two principles will be in conflict when the newer information is also the most important. If X is newer but more important information than Y, then given-before-new predicts that X should follow Y, while first-things-first says that X should precede Y. Most scholars who have addressed these issues have concentrated on some version of given-before-new, and we will likewise concentrate on this claim.

[3] This three-way distinction will later be collapsed into two categories, given/inferable and new, because of a small number of inferable NPs in the corpus analysis.

GIVENNESS, it allows us to use parallel terms for the two dimensions under study here, heaviness and newness.

1.3. FUNCTIONS OF CONSTITUENT ORDERING: SPEAKER-ORIENTED VS. LISTENER-ORIENTED ACCOUNTS. Why do speakers choose one constituent ordering over another? What functions does ordering serve? Past treatments of this issue have frequently looked at how ordering can facilitate communication. But which communicative processes might it facilitate? One way of looking at ordering is in terms of whether it benefits the speaker or the listener. One possibility is that it reflects the real-time constraints on producing an utterance, and is a strategy used mostly to take care of the speaker's needs. Another possibility is that speakers choose constituent orders to facilitate the process of comprehension for the listener. A third possibility, of course, is that ordering serves both functions, and that both motivations come into play at one time or another.

These questions tap into a larger issue about the degree to which speakers can maintain a model of the listener's needs, and design their utterances accordingly. Clark (1996, Clark & Marshall 1981) argues that, to a certain extent, speakers tailor their utterances to the current situation and the listener's needs, in the interests of successful communication. But speakers may not have the resources to take the listener's needs into account constantly (e.g. Horton & Keysar 1996). Anderson et al. (1997) for example, found that speakers' clarity of pronunciation reflected listener needs at a gross level, but not at a fine-grained level.

Thus, the functions of constituent ordering may be to facilitate the processes of production, comprehension, or both; we will review the treatments of this issue in the literature on constituent ordering.

FUNCTIONS OF END WEIGHT. Many researchers (e.g. Bever 1970, Kimball 1973, Frazier & Fodor 1978, Hawkins 1990, 1994, Gibson 1997) have sought to explain end weight in terms of the architecture of the human parser, thus focusing on its advantages for the listener. Bever (1970) sums up the basic idea: 'Save the hardest for last'.

Hawkins proposed a particularly explicit version of this idea, the principle of EARLY IMMEDIATE CONSTITUENTS.

> Constituents occur in the orders that they do so that syntactic groupings and their immediate constituents (ICs) can be recognized and produced as rapidly and efficiently as possible in language performance. Different orderings of elements result in more or less rapid IC recognition. (Hawkins 1994:57)

Hawkins assumes a top-down parser, whose first task in analyzing a VP will be to determine its immediate constituents—that is, the labels of its daughter nodes. He calculates how early the hearer will be able to identify all immediate constituents by counting the number of words it takes until all ICs are recognizable. This system predicts that 5 will be easier to comprehend than 4, requiring only four words to find all ICs instead of eleven.

(4) I $_{VP}$[gave $_{NP}$[the valuable book that was extremely difficult to find] $_{PP}$[to Mary]]

1 2 3 4 5 6 7 8 9 10 11

(5) I $_{VP}$[gave $_{PP}$[to Mary] $_{NP}$[the valuable book that was extremely difficult to find]]

1 2 3 4

Many real examples of HNPS provide anecdotal support for this sort of analysis.

For example, the other possible order for 6, given in 7, is almost unprocessable (brackets added).

(6) The Informativeness account shares
[with the processing account $_{PP}$]
[the claim that a subjectless sentence is not a grammatical option for the child, and that the omission is due to some aspect of performance $_{NP}$].
(Nina Hyams and Kenneth Wexler, 'Grammatical basis of null subjects in child language')

(7) The Informativeness account shares
[the claim that a subjectless sentence is not a grammatical option for the child, and that the omission is due to some aspect of performance $_{NP}$]
[with the processing account $_{PP}$].

A second possible function of end weight is to facilitate planning and production. Speakers construct utterances on the fly, frequently beginning sentences before the exact wording has been fully worked out (Clark & Wasow 1998). Postponing elements that are hard to produce, such as long and complex constituents, gives the speaker more time to formulate them. Sentences like 8, uttered frequently by baseball play-by-play announcers, are an example of HNPS in which planning seems like the obvious functional explanation.

(8) That brings to the plate Barry Bonds.

Ex. 8 is no easier to parse than *That brings Barry Bonds to the plate*. Indeed, the latter is slightly preferred by Hawkins's metric for evaluating alternative orderings. But the ordering in 8 provides the announcer with about an extra second to check the scorecard to see who is next at bat.

Speakers may have difficulty formulating any part of their utterances, for a variety of reasons; one source being the length and complexity of one constituent. When formulation is difficult, choices in constituent ordering allow speakers to postpone the long, difficult constituent while they utter the shorter, easier one.

This argument was made by Wasow (1997a,b), who examined examples of HNPS and DA constructions where the speaker's and listener's needs are in conflict, and found that speakers favored orders that benefited themselves rather than their addressees. A production account of ordering has been offered by Stallings et al. (1998), who used a production experiment to demonstrate the influence of end weight and lexical choice on HNPS. Similarly, Ferreira (1996) showed that production is facilitated when a construction admits choices in constituent order.

FUNCTIONS OF GIVEN BEFORE NEW. There are also clear advantages of given-before-new, from the perspectives of both the speaker and the listener. On one hand, beginning a sentence with reference to something that has been mentioned before provides a link between what has already been said and what is about to be said. This lends continuity to the discourse, which would facilitate comprehension for the listener. Furthermore, by virtue of having certain sentence positions recognized as the loci for new information, speakers can convey to their interlocutors just what they are saying that is new. While it may not affect the truth conditions of sentences, the choice of constituent ordering may have a considerable influence on what listeners take the point of an utterance to be. Thus, given-before-new facilitates efficient communication. This is illustrated in 9, in which the two italicized NPs are in the 'shifted' position at the end of their VPs, despite their lack of syntactic complexity.

(9) to take as input *a grammar* and to associate with it, as output, *an optimal structure* (Paul Smolensky, 1994 CUNY conference)

In this example, HNPS highlights the newness and importance of the NPs in question, as well as their parallel roles in the two conjuncts.

The tendency to postpone NPs that refer to newly introduced entities also has clear advantages for the speaker. While the processes that are involved in planning and producing speech are not fully understood, the production of an utterance minimally requires the speaker to: (a) decide what to say, (b) formulate the linguistic expressions to communicate that idea, and (c) articulate the utterance (Levelt 1989). All of these processes are usually accomplished in a very short time span during speech production. Given the time constraints of spontaneous conversation, many of these processes are likely to be more difficult when the speaker refers to something new, as opposed to something given: accessing the desired lexical items or choosing an appropriate linguistic expression may take more time; generating the phonetic plan is more difficult if the expression has not already been primed, and the speaker is likely to have more trouble in the articulation processes as well (Bock 1986, 1987, Bock & Irwin 1980, Tomlin 1998). Thus, the expression of new NPs is potentially more treacherous than referring to given NPs for which both conceptual and linguistic representations are already active. By postponing new, difficult NPs, speakers afford themselves extra time to complete these processes.

Ambiguity avoidance. Constituent ordering may also help the speaker to avoid potential attachment ambiguities, a strategy that would benefit the listener. Consider sequences of the form V + NP + PP. Except where the NP is a definite pronoun or proper name, such sequences present the parser with a potential ambiguity, for the PP may be attached to the VP or to the NP. If the same elements are presented in the order V + PP + NP, however, no such ambiguity exists. This is illustrated in 10.

(10) If you want to identify in the inner city the fifth graders who could really take off. . . (Patrick Suppes, in a talk)

If the phrase *in the inner city* occurred in its canonical position after the object NP, *the fifth graders who could really take off*, listeners would tend to interpret the PP as part of that NP. A similar attachment ambiguity for the adverbial *last July* in 11 is avoided by the HNPS construction.

(11) When Kennedy announced last July his plans to quit . . . (San Jose Mercury 1992)

Other manifestations of end weight, including DA, can likewise eliminate attachment ambiguities.

Since ambiguous syntactic structures put an extra burden on the parser, minimizing such ambiguities makes the hearer's task easier. Using constituent ordering for disambiguation requires the speaker to consider how a particular utterance would be understood. In this sense, ambiguity avoidance can be considered a hearer-oriented strategy.

However, potential attachment ambiguities are extremely common, suggesting that this is probably a relatively weak determinant of constituent order. A possible explanation for its weakness is that in most cases of attachment ambiguities, only one structure corresponds to a meaning that is pragmatically plausible. Ferreira and Dell (1998) investigated another type of variation, the presence or absence of *that* at the beginning of sentential complements, and found that the speaker's choice of wording was NOT motivated by an ambiguity avoidance strategy. By contrast, Schafer and Speer (1999) found that speakers used prosody to disambiguate potentially ambiguous utterances, in a natural communicative setting. Further research is needed to establish the extent to which a disambiguation strategy influences production, and whether it influences constituent ordering.

In sum, the many functional explanations for constituent ordering can be categorized in terms of facilitating comprehension processes, production processes, or both. Past research on constituent ordering has focused primarily on comprehension processes. In our study, we use a production paradigm to reveal the effects of newness and heaviness. This paradigm additionally provides data on disfluencies, which we use to investigate the effects of production difficulty on constituent ordering. Following Wasow 1997a,b, we propose that the postponing of heavy and new elements facilitates planning and production, in addition to the advantages it may provide for the listener.

1.4. HEAVINESS VS. NEWNESS. Many scholars have claimed that constituent ordering is dependent on either complexity or information structure, but few have tried to compare the two. Making such a comparison is complicated by the high correlation between them. Items which have been recently mentioned tend to be accessible to both speaker and hearer, and require less complex descriptions than items which are not (see, for example, Ariel 1990, Arnold 1998, Givón 1983). Therefore, items that are new to the discourse tend to be complex, and items that are given tend to be simple. The clearest example of this is the use of anaphoric pronouns, which refer to given information with a single (usually monosyllabic) word. Niv (1992), noting the correlation between newness and heaviness, goes so far as to speculate that all effects attributed to end weight can be explained in terms of information structure.

Two studies that explicitly compare complexity with information status are Siewierska 1993 and Hawkins 1994. Both employ Hawkins's theory for their metric of complexity and Givón's (1988) measures of 'topicality' to operationalize the notion of givenness. Hawkins concludes that his data, drawn from texts in a variety of languages, show that 'neither Givón's theory . . . nor the Prague School's . . . can be considered significant generalizations about discourse organization . . . Hence, there is no currently valid descriptive generalization for linear ordering in pragmatic terms' (237). Hawkins is not using the term significant in the technical sense of statistical significance. In fact, he does not report any significance tests. Furthermore, his conclusion is based on the fact that his structural theory makes correct word-order predictions on his data more frequently than the information-based measures. No analysis is reported in which one type of factor is held constant and the effects of the other tested. In short, Hawkins does not address the possibility that both structural and informational factors influence ordering. Siewierska (1993) uses the same two theories to analyze Polish texts and compare them as predictors of constituent order. Unlike Hawkins, she does test for statistical significance and compares cases where the two criteria make opposite predictions. She ends somewhat equivocally, saying that while informational factors seem to dominate structural ones in her study, more research is needed to determine their interactions.

In sum, despite the large literatures on the roles of both newness and heaviness in constituent ordering, few researchers have taken both into account simultaneously. Moreover, what little research there is that considers both is far from conclusive. In the following sections, we present new evidence, from both a corpus analysis and an elicitation experiment, showing that grammatical structure and informational factors both influence ordering.[4]

[4] There are also other factors that influence constituent ordering, a point to which we return at the end of the paper.

2. Corpus analysis

2.1. Methods and materials. This study looked at two constructions: heavy NP shift and the dative alternation. The corpus used was the Aligned-Hansard corpus, which consists of transcriptions of debates in the Canadian parliament. This corpus has two advantages. First, its sentences were originally spoken, although some sections were probably read from prepared texts;[5] hence, it provides a mix of speech and writing. Second, the corpus is divided into agenda items, each of which can be reasonably treated as a separate discourse. This facilitates the classification of NP tokens as referring to given or new entities.

The question we aimed to answer was whether grammatical structure and discourse status both affect constituent order, or whether one factor can account for all the variation. Toward this end, we first needed to extract examples of HNPS and DA. To investigate HNPS, we searched for examples of the (possibly discontinuous) strings *bring . . . to . . .* and *take . . . into account. . . .* These strings were chosen because Wasow (1997a) found that they are common sequences that occur frequently in both the canonical V + NP + PP order and in the 'shifted' V + PP + NP order. Idioms like *take into account* exhibit a far higher rate of the shifted order than nonidiomatic expressions with the same syntactic structure. The *bring to* data include both idiomatic and nonidiomatic uses. To investigate DA, the verb *give* was chosen as a representative. *Give* is by far the most frequently occurring verb that exhibits DA. Moreover, the subtle semantic differences between the prepositional and double object constructions are largely neutralized with *give* (see Goldberg 1995).

Using a string matching program (namely, the UNIX utility grep), we extracted all sentences containing occurrences of the relevant strings (including all verbal inflections) from a subset of the corpus. We then went through these one by one, throwing out those that did not exemplify the constructions under consideration in this study, e.g. uses of *bring . . . to . . .* in which *to* is an infinitive marker, not a preposition, and *give* in verb-particle constructions, such as *give . . . away*. The remaining utterances included 223 examples of *bring . . . to . . .*, 167 examples of *take . . . into account*, and 269 examples of *give*.

NO. OF EXAMPLES	CONSTRUCTION TYPE	EXAMPLE
223	*bring . . to . .*	It is enough to bring tears to the eyes of the Statue of Justice that stands silently in front of the Supreme Court of Canada.
167	*take . . into account*	It takes into account the need not to be overtaken by the constant evolution of the labour market.
269	*give*	The bank was told it should give its business to a friend of the Government.

Table 1. Examples from the corpus analysis.

We hand coded these utterances for constituent order, assigning values we will refer to as NONSHIFTED (V + NP + PP) and SHIFTED (V + PP + NP) for HNPS, and PREPOSITIONAL (V + NP + PP) and DOUBLE OBJECT (V + NP + NP) for DA. We also coded for heaviness and newness. Heaviness was measured as the relative length of the two constituents, in terms of number of words (number of words in the theme NP minus the number of words in the goal NP). Because of the small number of tokens in some cells, each item was put into one of three categories of relative length. The

[5] In addition, the absence of disfluencies (e.g. *um*, *uh*, repeats, repairs) suggests that the transcript is edited.

a) th < go	theme NP length − goal NP length =	−2 or less
b) th (=) go	theme NP length − goal NP length =	between −1 and 1
c) th > go	theme NP length − goal NP length =	2 or more

TABLE 2. Categories of heaviness for dative alternation.

categories of heaviness for DA are shown in Table 2. For HNPS, which had a larger number of tokens, we were able to use a finer-grained categorization of heaviness, shown in Table 3.

a)	DO ≪ PP	DO length − PP length =	−4 or less
b)	DO < PP	DO length − PP length =	between −1 and −3
c)	DO = PP	DO length − PP length =	0
d)	DO > PP	DO length − PP length =	between 1 and 3
e)	DO ≫ PP	DO length − PP length =	4 or more

TABLE 3. Categories of heaviness for heavy-NP-shift.

Newness was coded as one of three categories: given, inferable, and new. These were identified according to whether the referent had previously appeared within the current agenda item. The number of NPs coded as inferable was too small to be informative, so, for purposes of our analyses, inferables were classified as given information.[6] For the DA sample we coded the newness of both constituents. For the HNPS samples, however, we coded newness only for the direct object, since newness is not applicable to prepositional phrases like *into account*.

2.2. RESULTS. The purpose of this study was to investigate the effects of newness and heaviness on constituent ordering in DA and HNPS. For this reason, we looked at the proportion of uses of each type of constituent order under the different conditions of newness and heaviness. We analyzed the relationship of each factor to constituent ordering by submitting our data to stepwise logistic regressions, using SPSS 6.1. To test the reliability of patterns observed for each factor and potential interactions, we used SPSS to calculate the log likelihood ratio for each factor, comparing the model WITH that factor to a model WITHOUT that factor. This test is similar to VARBRUL (commonly used in sociolinguistic studies), except that it allows us to investigate the interactions among independent variables. For each data set we ran both step-up and step-down analyses; in all cases they were identical.

The main finding of the corpus analysis is that both heaviness and newness are significantly correlated with constituent ordering, and that neither correlation can be reduced to the other. This holds for both dative alternation (Fig. 1) and heavy NP shift (Fig. 2).

Figure 1 shows that dative alternation correlated with both newness and heaviness, such that speakers used the double object construction more when (a) the theme was newer than the goal, and (b) when the theme was heavier than the goal. Both newness (-2 * Log LR $= 7$, $p < .01$) and heaviness (-2 * Log LR $= 99.5$, $p < .001$) entered the logistic regression, but the interaction newness × heaviness did not.

Figure 2 shows the results for the heavy-NP-shift corpus, in which shifted orders are more frequent when the DO is new, and when the DO is relatively longer than the PP. This graph includes data from both the *bring to* and *take into account* corpora. The rate of shifting in this sample was influenced by heaviness, newness, and verb (*take*

[6] The results of this study were not substantially changed if the inferable NPs were grouped with the new ones or omitted altogether.

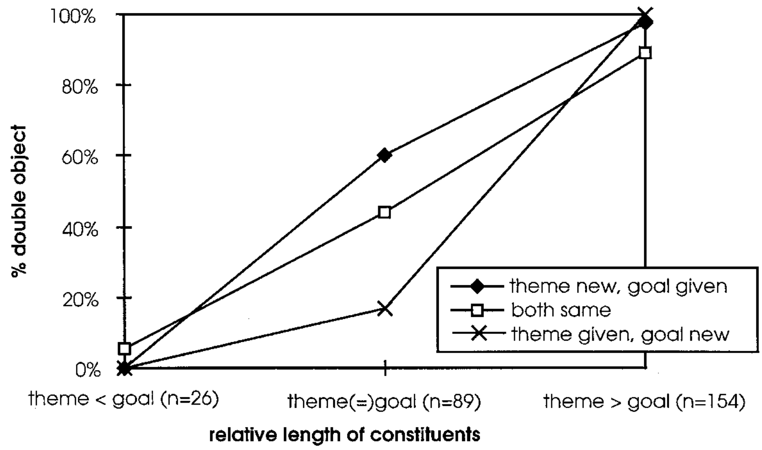

FIGURE 1. *Give* corpus analysis results.

into account vs. *bring to*): the logistic regression selected newness (-2 * Log LR $=$ 23.5, p $<$ .001), heaviness (-2 * Log LR $=$ 120.9, $p < .001$), and verb (-2 * Log LR $=$ 31.4, p $<$ .001) as significant factors, but did not select any of the interactions (newness × heaviness, newness × verb, heaviness × verb, newness × heaviness × verb).

Figures 3 and 4 show that when the *bring to* and *take into account* items are considered separately, the same results hold. *Take into account* has an overall higher rate of shifting than *bring to*, but separate logistic regressions for each verb sample selected both newness and heaviness as significant factors (p's $<$ .01), but not the interaction newness × heaviness.

2.3. DISCUSSION. The corpus analyses showed that both DA and HNPS are influenced by the relative newness and heaviness of the two constituents. In both cases, speakers tended to produce the relatively newer and heavier constituent later. For DA, the use of double object constructions was higher when the theme was newer and heavier than

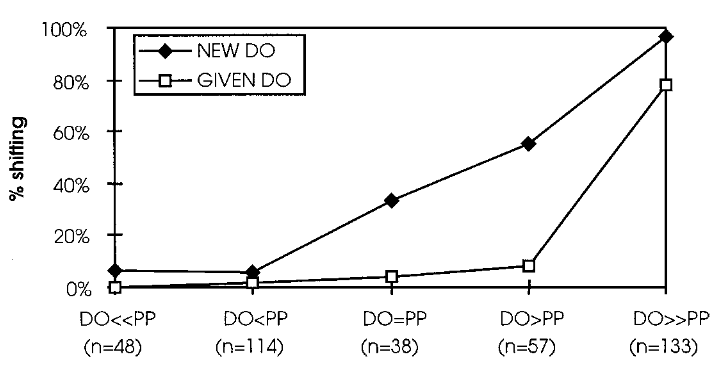

FIGURE 2. Heavy-NP-shift corpus analysis results.

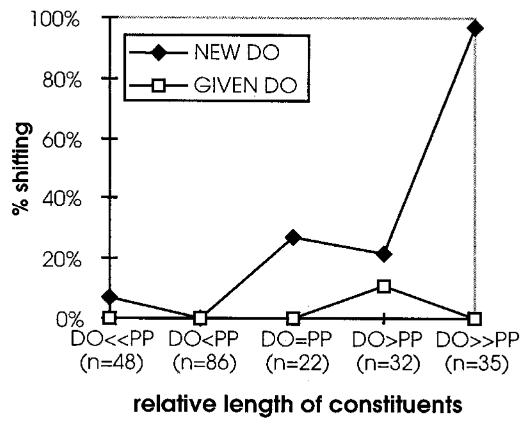

FIGURE 3. *Bring to* analysis results.

the goal. For HNPS, shifting increased when the DO was newer and heavier than the PP. These data clearly show that variation in word order cannot be accounted for by just newness or heaviness alone.

There are three further observations we make about these data. First, in both cases, heaviness accounted for more of the variation than newness, though we do not take this as evidence of a general primacy of heaviness over newness. See §5 for a discussion of this point.

A second observation we can make about Figs. 1 and 2 is that the effect of newness appears to be greater for the middle categories of heaviness; that is, newness has more of an effect when heaviness does not make strong predictions either way. When one constituent is considerably longer than the other, newness does not account for much of the data, but when the two constituents are closer in length, newness effects are more clearly observable. However, we are extremely cautious in this conclusion, because the interaction newness $\times$ heaviness does not contribute significantly to the logistic regression models. We will discuss this issue further below.

Third, there was a reliably greater rate of shifting for the *take into account* corpus than the *bring to* corpus. We expected *take into account* to have a relatively high rate

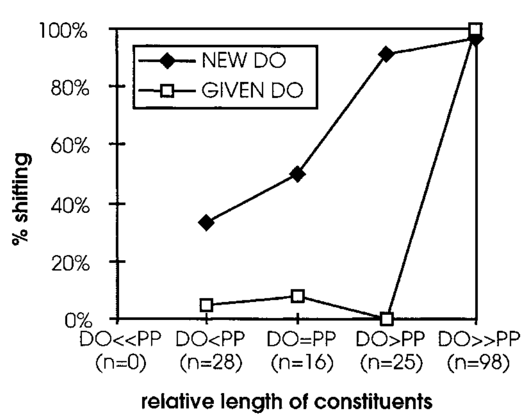

FIGURE 4. *Take into account* analysis.

of shifting in part because the prepositional phrase is extremely short (two words), but the results show that its higher rate of shifting holds for each category of heaviness, suggesting that something about the construction itself is leading speakers to shift more for *take into account* than *bring to*. One explanation draws on the fact that collocations (fixed expressions) have a higher rate of shifting than noncollocations (Wasow 1997a). Wasow further shows that semantically opaque collocations, like *take into account*, have a higher rate of shifting than other collocations, for example, *bring to an end*. The degree of semantic transparency is what differentiates idioms from other collocations (Nunberg et al. 1994). Although some uses of *bring to* are idiomatic (like *bring to bear*), others aren't. As Wasow (1997a) points out, once a speaker has decided to use an idiom like *take into account*, it is easier to produce it all together, leading to a shifted structure.

Thus the corpus analyses provide evidence that both newness and heaviness influence constituent ordering, but also point to a role for other factors, such as the degree to which the verb and prepositional phrase constitute a fixed expression. We now present a rather different type of study, investigating the same questions in an experimentally controlled informal speech situation.

3. *Give* EXPERIMENT. The second method we used to investigate these issues was an elicitation experiment. This method focused only on constituent order in DA, and again used *give* as the verb. We chose to focus on DA, because this alternation exhibits a much higher degree of variation than HNPS, in which the unshifted ordering is used over 90% of the time (Wasow 1997a). In a pilot experiment we also found that it was easier to elicit variation in DA than HNPS (Arnold & Wasow 1996). We hypothesized that it would be easier to observe the effects of discourse status and syntactic complexity in DA because neither form is particularly dominant.

A persistent difficulty in experimental psycholinguistics is that most easily controllable experimental paradigms require rather unnatural ways of using language. But studying speech in a natural setting means that the experimenter loses the ability to control the environment, and many factors can get confounded, making it difficult to interpret the outcome. This study was designed to allow us to control the environment, but still elicit language production in a relatively natural situation.

3.1. METHODS AND MATERIALS. Participants, who worked in pairs, were presented with three sets of three animals each, such as those shown in Fig. 5, and over a hundred objects. Participants were instructed that they would be working together, giving each object to different animals. One participant (whom we will call THE DIRECTOR) was assigned the role of giving instructions; the other (whom we will call THE ACTOR) was given the role of carrying out those instructions by manipulating the objects. The director received a set of cue cards, indicating which object was to be given to which character. Crucially, the cue cards used pictures, rather than linguistic instructions, to indicate which action should be taken. Actors also received a set of cue cards, containing prompts for the director to give an instruction. The reason for this will be discussed below.

Participants were assigned to one of three groups, each of which had identical cue cards, but in different orders, and using different animals. Group 1 used ducks, squirrels, and horses, group 2 used cats, dogs, and frogs, and group 3 used ducks, rabbits, and giraffes. There were 24 pairs of participants, each of which performed the experiment twice, switching roles the second time. This resulted in a total of 48 sessions, where a session is one time through the cards.

We were trying to learn what word order the participants would produce, and how

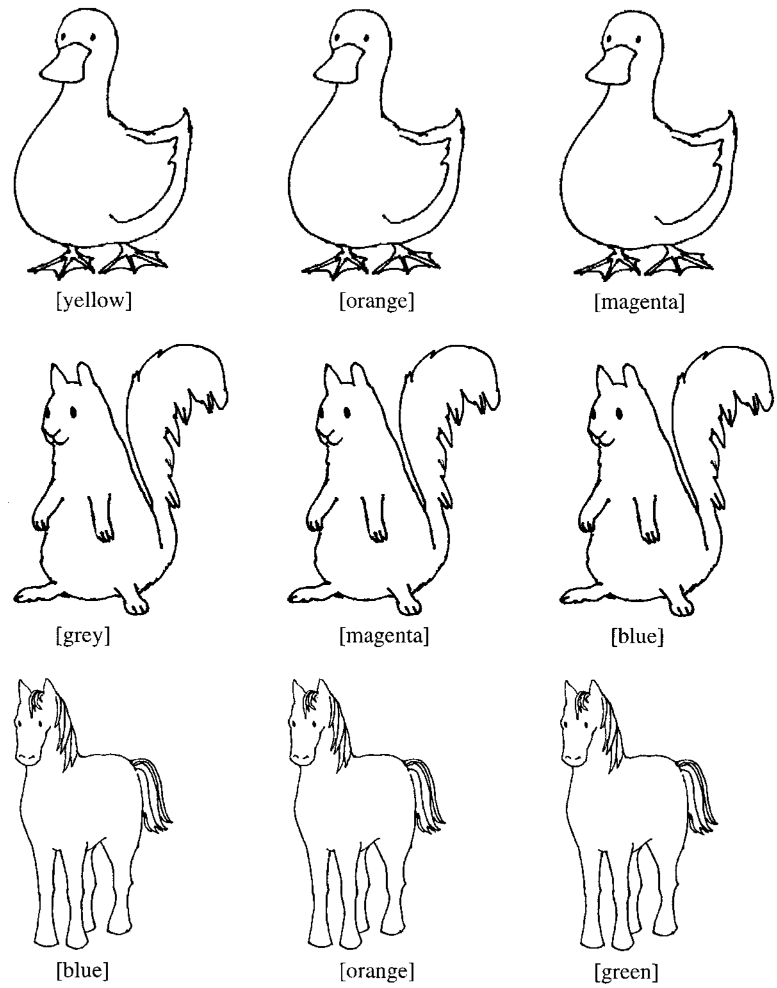

FIGURE 5. Example characters from the experiment.
Note: the color of each character is indicated below each picture. In the experiment, the materials were in color and purely graphic, with no words written on them.

this would be affected by the discourse status of the referent and the complexity of the NP used to denote it. We were interested in studying a natural speech situation, but still wanted to manipulate the environment to make people produce a large number of DA constructions. For this reason, we put people in a situation where they would be issuing instructions about giving objects to characters. But if the situation were to consist solely of acts of giving, there would be no reason for the participants to use the word *give*, since it would become redundant after the first few trials. Therefore, the experiment included 24 trials involving giving actions, and 12 fillers consisting of taking away or exchanging items. There were also three practice cards at the beginning of each session. An example set of *give* instruction cards is shown in Fig. 6, and example filler cards are shown in Fig. 7.

An essential feature of the experiment was that we were able to manipulate the discourse status and complexity of participants' reference to objects and characters. To control the newness of the referents, we used the actor's cue cards. Each trial was

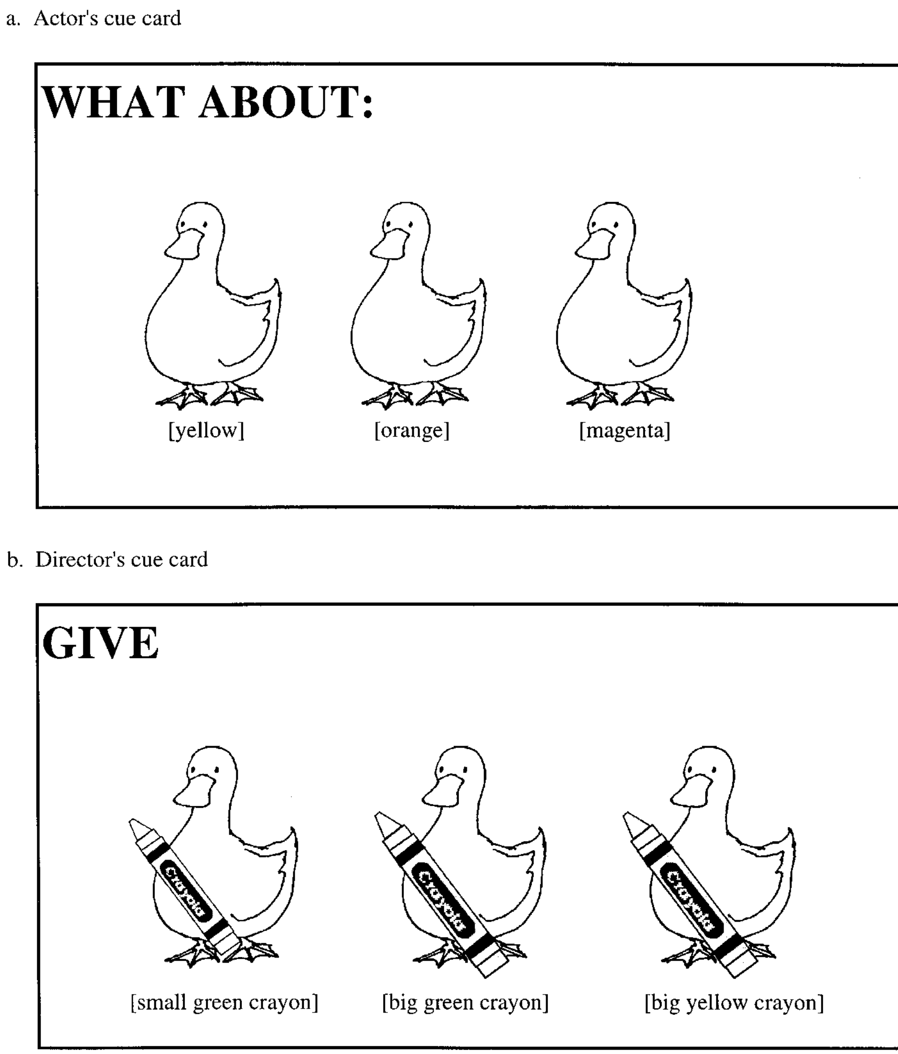

FIGURE 6. Sample set of cue cards. Characters and objects in the actual experimental materials were all in color and purely graphic.

initiated by the actor asking a question, which established either the objects or the characters as given. For example, if the actor's cue card showed a picture of the three ducks, as in Fig. 6a, he would ask 'What about the yellow duck, the orange duck, and the purple duck?', thus establishing the ducks as given. The director would then follow her own cue cards in order to know which objects to tell the actor to give to which characters.

The heaviness of the expressions participants used was manipulated by means of contrast sets. The primary manipulation of heaviness occurred with the objects, which belonged to one of two groups: simple or complex. The complex items belonged to a contrast set of four or six items, which differed on two qualities. The crayons in Fig. 6b are examples of complex items; they differ from each other in both thickness and color. In order to identify each object uniquely, the director needed to use a fairly long NP, such as *the small green crayon*, or *the large yellow crayon*. The simple items, in

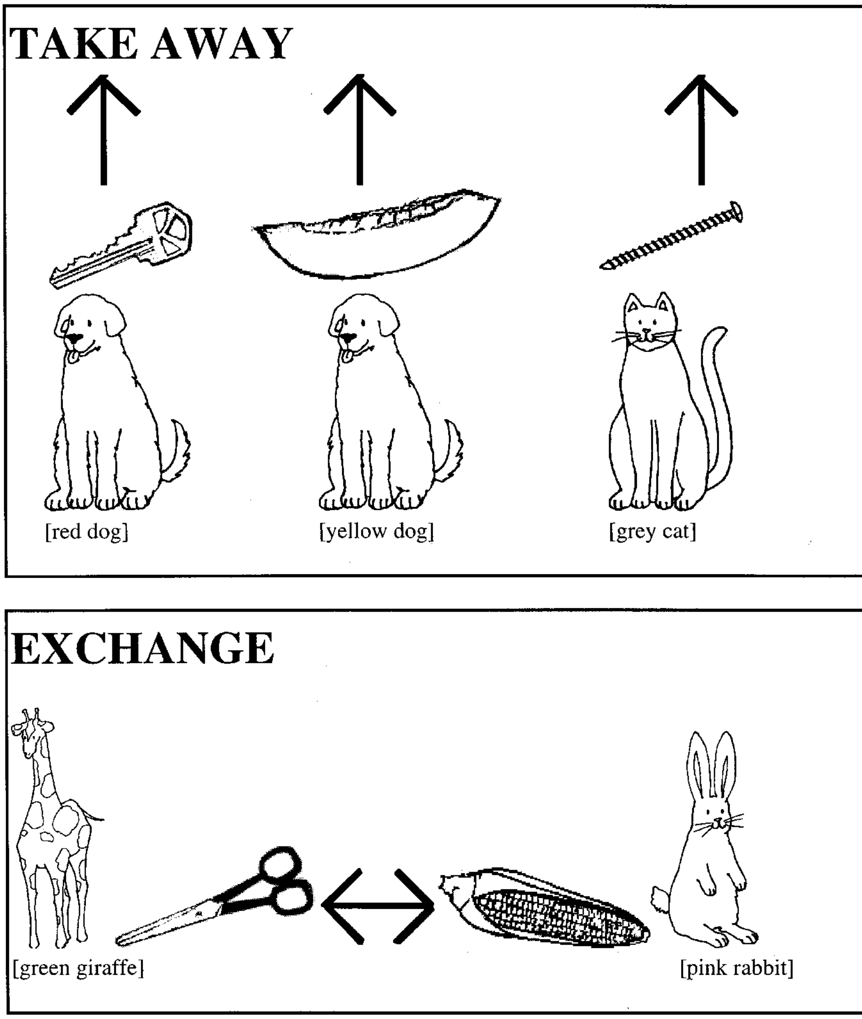

FIGURE 7. Sample Filler Cards. Characters and objects in the actual experimental materials were all in color and purely graphic.

contrast, could be identified with a two-word NP, such as *the key*, or *the scissors*, in Fig. 7. In contrast with the items, the characters were all of uniform complexity, falling in between the simple and complex objects. The characters all belonged to sets of three, which differed on one feature, color. *Give* instructions always included three characters of the same type (e.g. ducks), so that in order to uniquely identify any single character, the participants would most likely use a three-word NP, such as *the yellow duck.*

The only restriction was that the actors were required to ask about all three objects at the same time. This constraint was imposed to avoid the use of pronouns, as might happen in a hypothetical exchange such as 12.

(12) A: What about the yellow duck?
B: Give it the small green crayon.
A: What about the. . . .

Apart from this, there were no restrictions placed on the participants; they were involved in a fairly natural speech situation.

The entire experiment was tape-recorded and transcribed. Each instruction was coded for construction type, newness, and heaviness. Participants used many different types of constructions, not only prepositional or double-object constructions. These constructions were grouped according to whether the goal or the theme argument came first. The range of uses is listed in Table 4.

CONSTRUCTION	EXAMPLE
GOAL FIRST	
• double object	Give the white rabbit the carrot.
• heavy-NP-shifted	Give to the white rabbit the carrot.
• passivized goal	The white rabbit should be given the carrot.
• double object, goal topicalized	The white rabbit, give the carrot.
• prepositional, goal topicalized	To the white rabbit, give the carrot.
THEME FIRST	
• prepositional	Give the carrot to the white rabbit.
• passivized theme	The carrot should be given to the white rabbit.
• prepositional, theme topicalized	The carrot, give to the white rabbit.

TABLE 4. Constructions used in the dative alternation experiment.

Since newness was manipulated by establishing either the theme or the goal as given, this variable was coded as either ‘theme given’, or ‘goal given’. In most cases this was established by means of a question, but in some cases the director mentioned an object or character immediately before giving the instruction, as in 13.

(13) Object-mover: What about the red dog, the yellow dog, and the purple dog?
Director: Um, you should look for the tic tacs and find the box with the white and green tic tacs and give that to the red dog.

In this case, the object or character mentioned by the director was coded as ‘given’. In some cases, the director mentioned both the character and object, or the categories of characters and objects, as in 14.[7]

(14) You are going to give the ducks objects. You are going to give the yellow duck A key. ’Kay. You’re going to give a du- the red duck, um, a melon. And you’re going to give THE um pink duck A um rectangular eraser, it’s black.

If this happened, the item was coded as ‘both given’, resulting in three levels for this variable.

Heaviness was coded as the relative length of the two constituents, in three categories: (a) theme length shorter than goal length, (b) theme length equal to goal length, and (c) theme length longer than goal length. In addition, each instruction was coded for the type and location of disfluencies, the purpose of which will be discussed below.

3.2. PARTICIPANTS. Forty-eight people from the Stanford community, ages 18–25, participated in this experiment for $7 each. The time to complete the experiment was approximately one hour. All participants were native speakers of English, defined as having learned English before age five.

[7] In these transcriptions, THE is used to indicate when the speaker pronounced *the* as [ðij]. When the vowel was lengthened, these cases are transcribed as THEE [ði:j] or thee [ðə:]. Similarly, the determiner *a* was transcribed as A when the pronunciation was [ej], and a for [ə]. These and other disfluencies will be discussed in §4.

3.3. RESULTS. Because this experiment involved a naturalistic speech situation, participants were completely unrestricted in how they worded their instructions. This meant that in many cases they did not use the word *give*, instead choosing expressions like *The ball goes to the cat*, or *The cat gets the ball*. We were only interested in what happened with the alternating dative *give*, so all other instructions were excluded from the analysis. Of course, the same lack of coercion about wording guaranteed that when participants DID use *give*, it was in a natural way.

Out of the 24 pairs of participants, one pair was thrown out because they articulated the dependent variable of the experiment, during the exchange shown in 15.

(15) Actor: [to partner] mmm—what about,
[to Experimenter] wait, can we talk about it . . .
Experimenter: sure
Actor: [to partner] okay, will you say the object first, in any situation, and then I can find it better?
Director: okay.

In addition, four sessions were thrown out because the director exhibited no variation in the *give* commands, producing one hundred percent prepositional constructions. A total of 42 sessions were included in the analysis. In the results reported here, we also excluded cases where the instruction was elliptical, as in the second and third instructions in 16.

(16) Example of elliptical constructions
Give the pink giraffe the plum, *the green giraffe the pear . . . and the yellow giraffe the tur er uhh plum.*

Elliptical utterances were excluded because the word order in these constructions is highly constrained to match the word order of the initial construction, and was not expected to be as sensitive to the variables of interest. This resulted in a total of 1684 instructions that were included in the final analysis.

Figure 8 shows that both newness and heaviness correlated with participants' choice of word order: goal-early constructions were produced more often when the theme was newer than the goal and when the theme was longer than the goal. These data were submitted to two logistic regression analyses, both of which included the factors newness, heaviness, production difficulty (to be explained below), and the interaction new-

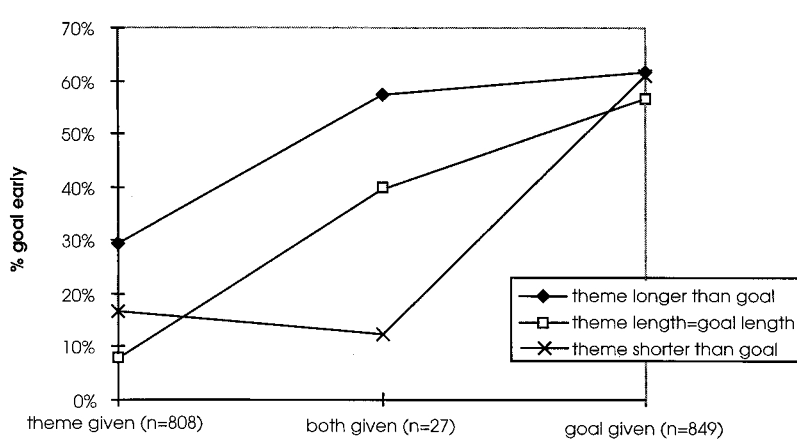

FIGURE 8. Dative alternation experiment results.

ness $\times$ heaviness. One analysis also included participant identity as an additional independent variable, and the other included item number. The inclusion of these factors in the logistic regressions adapts Clark's (1973) proposal that both participants and items should be included (one at a time) as random variables in analyses of variance.

In both analyses, the model selected newness ($-2 * \text{Log LR} = 298$, $p < .001$), heaviness ($-2 * \text{Log LR} = 9.5$, $p < .005$), production difficulty ($-2 * \text{Log LR} = 9$, $p < .005$), and the interaction newness x heaviness ($-2 * \text{Log LR} = 20$, $p < .001$).[8] Participant identity was also selected as significant ($-2 * \text{Log LR} = 671$, $p < .001$), but item number was not.

3.4. DISCUSSION. As in the corpus analyses, the experiment provided clear evidence that both newness and heaviness have significant effects on constituent ordering. People tended to produce their utterances such that the relatively newer and heavier constituent came later in the sentence. That is, they followed patterns of both given-before-new and light-before-heavy.

The experimental data also displayed a significant interaction between newness and heaviness: heaviness had more effect on constituent order when both constituents were given than when they differed in newness. That is, weight effects were stronger when newness made no prediction about the outcome. This finding will be discussed further below.

The experiment also revealed a significant effect of participant identity. There was in fact enormous variation among participants, whose use of goal-early constructions ranged from 2% to 98%. (As mentioned above, there were also four participants who used 100% prepositional constructions; these participants were excluded from the analysis for lack of variation.) This result may indicate an individual preference for word order, such that a given individual may use one word order more over his or her lifetime, or it may mean that different people respond to conditions of newness and heaviness differently. Perhaps some people make their word order decisions more on the basis of other factors. In any case, however, participant identity was only one of the factors selected by the logistic regression. The fact that newness and heaviness were also selected means that they cannot be explained away in terms of personal preferences.

4. CONSTITUENT ORDER AS THE RESULT OF PRODUCTION CONSTRAINTS. Both the experiment and the corpus analyses support the conclusion that constituent ordering is influenced by both newness and heaviness. Why is this? What is the function of postponing new, heavy NPs? Here we return to the issue of production-centered vs. comprehension-centered accounts of constituent order. Many scholars have discussed end weight in terms of its effect on parsing (e.g. Bever 1970, Frazier & Fodor 1978, Hawkins 1990, 1994, Kimball 1973); others have noted the advantages that choices in constituent order bring to the speaker (e.g. Bock 1987, Ferreira 1996, Wasow 1997a,b).

One commonality between newness and heaviness is that the newer and heavier a constituent is the more difficult it is to produce. And when a constituent is difficult to produce, speakers can take advantage of syntactic flexibility to postpone the difficult constituent until later in the utterance (Bock 1987). Following this sort of strategy would afford the speaker an additional moment to plan and produce a difficult phrase.

[8] These statistics are from the analysis including participant identity as an independent variable. The results from the analysis by items were similar (newness: $-2 * \text{Log LR} = 246$, $p < .001$; heaviness: $-2 * \text{Log LR} = 7.5$, $p < .007$; production difficulty: $-2 * \text{Log LR} = 9.6$, $p < .005$; newness $\times$ heaviness: $-2 * \text{Log LR} = 16$, $p < .001$).

New and heavy constituents are likely to be harder to produce than given, light ones, for many reasons. Recall that language production involves three major stages: (a) conceptualizing a message, (b) formulating the grammatical characteristics of the message, and (c) articulating it (Levelt 1989). The formulation stage involves selecting the appropriate lexical items, retrieving their phonological and morphological forms, choosing an appropriate syntactic frame, and reconciling the lexical and syntactic segments. All of these processes occur extremely rapidly, as social constraints force speakers to contribute to the situation at hand (Clark 1996, Clark & Wasow 1998). Discourse status and syntactic weight could influence ease of production at any of these stages. Lexical retrieval and phonological processing should be more difficult for long, complex phrases than shorter ones, simply because they contain more linguistic material. Furthermore, complex phrases are normally conceptually more difficult than simpler ones, and the speaker may have difficulty deciding exactly which features of an entity to describe.

There is also evidence that given information is easier to produce than new information. When something has been recently mentioned (when it is given), all discourse participants can be expected to have activated representations of that entity at conceptual, lexical, and phonological levels. Therefore the subsequent production of a reference to that entity could be facilitated at any of the multiple stages of production. Experimental evidence has shown that speakers tend to put NPs early in the sentence when they refer to referents that are conceptually accessible (Bock & Irwin 1980, Bock & Warren 1985, Tomlin 1998) or lexically primed (Bock 1986, Bock & Irwin 1980). Bock (1987) also showed that word-order variation is sensitive to the accessibility of word representations at the phonological level: when the phonological representation of a word was inhibited through the priming of a phonologically related word, speakers tended to put the less accessible constituent later in the sentence.

Like these studies, our findings suggest that constraints on planning and production lead to patterns of given-before-new and light-before-heavy. But our findings could also be taken as evidence for an alternate view, that speakers postpone new, heavy constituents to make processing easier for the hearer. As Hawkins (1990, 1994) has argued, postponing heavy elements reduces memory load in parsing, thus making the hearer's task easier. Given information is also generally considered more accessible than new information (e.g. Ariel 1990, Arnold 1998, Chafe 1994, Gundel et al. 1993), so one could also explain the postponement of new information in terms of the speaker's concern for the needs of the hearer. Our findings that both newness and heaviness influence constituent order are not, in and of themselves, sufficient to choose between a speaker-centered or addressee-centered account of constituent ordering.

Our experiment does provide an additional source of evidence about utterance planning that can help test the hypothesis that it plays a role in choices of constituent ordering. That source of evidence is disfluencies, which are direct indicators of production difficulty. As argued by Clark and colleagues (e.g. Clark 1994, 1996, Fox Tree & Clark 1997), speakers need to contribute to a conversation in a timely fashion. When speakers are not able to meet the goal of fluent production, they resort to a variety of disfluencies in order to stall for time. These disfluencies may be fillers such as *um* or *uh*, or they may surface in the form of repeats (*the . . . the elephant*), repairs (*the eleph . . . the squirrel*), or metalinguistic comments (*the el*, *I mean*, *the squirrel*). In addition, Fox Tree and Clark (1997) have shown that pronouncing the vowel in *the* as [ij] rather than as a schwa is also a disfluency of sorts, in that speakers tend to use this pronunciation more when they are having difficulty.

Disfluencies very likely have a number of different underlying causes, corresponding to different stages of production: speakers may have trouble figuring out what they want to say, choosing the words to say it, generating a phonetic plan, or implementing it. For this reason, disfluencies are a blunt instrument in measuring production processes. But, they are DIRECT evidence of difficulty in production. While it seems natural that such factors as the heaviness of an NP or the newness of its referent might increase the difficulty of producing it, we don't know whether they do in every case. But when speakers produce disfluencies, we can be fairly certain that they are experiencing difficulty.

Although speakers produced disfluencies at many points of the utterance, the type we were most interested in was when they were disfluent at the beginning of the instruction. Research shows that speakers tend to do much of their utterance planning near the beginnings of clauses (Clark & Wasow 1998), so this is where we would expect planning problems to be most evident. Consequently, we coded the utterances from our experiment for whether the speaker had produced one or more disfluencies, of any type, during the initial portion of the utterance. The initial portion was defined according to the schema in Table 5.

CONSTRUCTION	INITIAL REGION	EXAMPLE
• double object	VERB	*Give* the white rabbit the carrot.
• heavy-NP-shifted	VERB	*Give* to the white rabbit the carrot.
• passivized goal	GOAL	*The white rabbit* should be given the carrot.
• double object, goal topicalized	GOAL	*The white rabbit,* give the carrot.
• prepositional, goal topicalized	GOAL	*To the white rabbit,* give the carrot.
• prepositional	VERB	*Give* the carrot to the white rabbit.
• passivized theme	THEME	*The carrot* should be given to the white rabbit.
• prepositional, theme topicalized	THEME	*The carrot,* give to the white rabbit.

TABLE 5. Coding schema for location of early disfluencies.

The results in Figures 9 and 10 show that when speakers were disfluent during the initial part of the utterance, they were significantly more likely to produce a construction with the goal early, rather than the theme early.[9] As Figs. 9 and 10 show, the effect of production difficulty (as indicated by the presence of disfluencies) appeared across all categories of newness and heaviness.

One might ask, however, whether disfluencies are correlated with the choice in

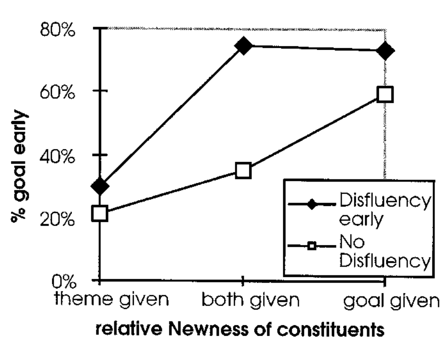

FIGURE 9. Production difficulty by newness.

[9] As a matter of fact, disfluencies in all regions correlated with an increased rate of constructions that put the goal early, but this correlation was highest for disfluencies in the first part of the utterance.

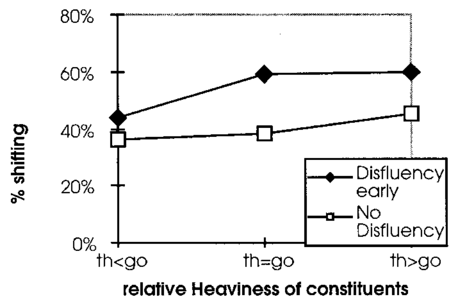

FIGURE 10. Production difficulty by heaviness.

construction because they are in fact highly correlated with the other variables of interest, newness and heaviness. For this reason, the factor PRODUCTION DIFFICULTY (presence vs. absence of disfluencies in initial region) was included in the logistic regression analysis, along with the other factors. As reported above, this factor was selected as significant ($p < .005$); that is, when the speaker was disfluent during the first part of the utterance, the tendency to put the goal early was significantly greater.

The strategy of starting out with the goal when you are disfluent is not surprising, if we consider the experimental design. There were only 9 characters, but over 100 objects. This means that not only was each character perceptually more salient than any particular object, but also that the participants mentioned the characters more times throughout the experiment, and were likely to become familiar with the characters and their denotations fairly quickly. This situation parallels that of the real world, where goals are often people, which are more likely than objects to be given in the discourse.[10] Therefore, when speakers were having difficulty producing the utterance, they were more likely to produce the easier, more familiar goal NP first, and then turn to the more difficult theme NP.

Thus, data on disfluencies in our experiment support Wasow's conclusion (1997a, b) that constituent-ordering preferences facilitate utterance planning. Our interpretation of these results is also more generally consistent with an incremental view of production, in which speakers do not always wait until their utterance is completely formulated before they start speaking (Clark 1996, Fox Tree & Clark 1997, Ferreira 1996).

5. SYSTEMS OF INTERACTING CONSTRAINTS. The results presented here suggest that multiple factors contribute to the choice of a particular grammatical form. The two we have investigated here are weight and discourse status. Using both a corpus analysis and a production experiment, we demonstrated that variation in constituent order is best understood when both factors are taken into account, and neither can be accounted for in terms of the other.

Our findings from the dative alternation experiment also revealed an interaction between newness and heaviness. Both factors played a significant role, but heaviness had the largest effect on utterances where both constituents were given. This interaction

[10] The idea that goals are often more salient than themes is supported by the data on pronoun usage in the DA corpus analysis, where 40% of goal NPs and 2% of theme NPs are pronominal. Since pronouns are used for highly salient referents, these data suggest that goal NPs are in general more salient than theme NPs.

is consistent with findings from diverse studies on language processing: when two factors are found to influence a particular choice or interpretation, the effect of each is usually stronger when the other factor is less constraining (see MacDonald 1996).

This interaction can also be observed in the differences between the results of the corpus analysis and the experiment. In the corpus analysis, heaviness played a larger and more consistent role, while in the experiment, newness had the biggest effect on the choice of construction. The reasons for this discrepancy are likely to stem from differences in the data sets and from the fact that our measurements of newness and heaviness were simplified operationalizations of the real psychological factors.

The first difference between the corpus analysis and the experiment lies in how newness was calculated. In both cases, this factor was operationalized as a binary feature, although we believe this is only a rough indicator of the accessibility of the referent, which is a graded property (Ariel 1990, Arnold 1998, Chafe 1994, Givón 1983, e.g.). However, our two studies made the binary distinction differently. In the corpus analysis, newness was calculated in terms of whether the referent had been previously mentioned or could be inferred from something that had been mentioned within the agenda item. In some cases, the previous mention may have occurred many sentences earlier. In the experiment, an element was established as given when the actor referred to it in a question about what to do, or when the director mentioned it immediately prior to the instruction. This meant that, in most cases, the item had been mentioned in the immediately preceding utterance, so was likely to be highly salient for both speaker and listener (see Arnold 1998 for a discussion). Thus, the elements categorized as ‘given’ in the experiment were generally more accessible than those categorized as ‘given/inferable’ in the corpus. This higher level of accessibility in the experiment may account for the finding of a stronger effect of discourse status in the experiment.

The second difference between the studies involves the relative heaviness of the two constituents. In both studies, we measured grammatical weight as the relative length of the two constituents. We further categorized these differences into a few discrete categories, because of the small number of tokens in some cells. However, 80% of the characters and objects in the experiment were described using NPs between one and seven words, so the average difference in constituent length was relatively small. In contrast, the naturally occurring data from the corpus analysis included a wide range of NP lengths. The differences between heaviness categories in the three data sets are shown in Table 6, which shows that there is a much bigger weight contrast between the two constituents in the corpus data than in the experimental data. Accordingly, heaviness was a much stronger determinant of word order in the corpus analysis than in the experiment.

Another difference in heaviness was that in the experiment, pronominal references

	HNPS CORPUS	DA CORPUS	DA EXPERIMENT
phrase 1 $\ll$ phrase 2[a]	−7.3		
phrase 1 < phrase 2	−1.5	−9.8	−1.1
phrase 1 = phrase 2	0	−0.2	0
phrase 1 > phrase 2	1.8	9.1	3.7
phrase 1 $\gg$ phrase 2	11.4		
RANGE	−21 to 44	−29 to 35	−8 to 20

TABLE 6. Average difference in length (# words phrase 1 − # words phrase 2) for each category of heaviness[b].

[a] For the dative alternation data, phrase 1 = theme NP, phrase 2 = goal NP. For the HNPS data, phrase 1 = direct object NP, phrase 2 = prepositional phrase.

[b] Categories of heaviness are defined slightly differently for each data set (definitions above).

were discouraged by the experimental design, and constituted only 2% of all the NPs. But pronouns are often used in real life: 21% of the DA corpus data were pronouns, as were 7% of the HNPS data. The use of pronouns is important because they are the lightest NPs in English. In fact, when the theme NP is an unstressed pronoun and the goal NP is not, the double-object construction is ungrammatical (**Give the man it*).[11] Although pronominal goals can occur grammatically in both constructions, they occur much more frequently in the double-object construction. Further, there appears to be a categorical restriction against heavy-NP-shifting when the direct object is a pronoun.

In sum, the NPs in the corpus analysis exhibited far more variability in heaviness than those in the experiment: in general, the heavy ones were heavier and the light ones were lighter in the corpus. In addition, as explained above, the given referents in the experiment were generally more salient than those in the corpus analysis. These facts, taken together, support the conclusion that heaviness and newness interact with each other. When one factor is less constraining, the other has more of a chance to influence the speaker's choice of constituent order. Heaviness was more constraining in the corpus analysis, so effects of newness were dampened; newness was more constraining in the experiment, so effects of heaviness were restricted.

The overall picture that emerges from the comparison of the two studies is that the role of each factor depends in part on the strength of competing factors. When there is a big weight difference between constituents, there is a strong tendency to produce the light argument early, and discourse status may not play as large a role. In contrast, when one argument is extremely accessible, by virtue of having been mentioned in the immediately preceding clause, discourse status will influence constituent ordering more than weight.

This evidence is consistent with a number of models that view language processing as a constraint-based system, where the strength of a constraint is greater when competing constraints are weak (e.g. Bates & MacWhinney 1989, MacDonald et al. 1994, Trueswell & Tanenhaus 1994). Our data suggest that variation in constituent ordering cannot be explained by a single factor, and that the speaker's choice may be driven by different considerations on different occasions, depending on which factors are strongest or which needs are most pressing.

Although we have focused only on the effects of discourse status and syntactic complexity, some items in our data sets are not accounted for by either factor. For example, 17 has a longer, newer constituent before a shorter, given one.

(17) Having given much consideration and attention to it . . .

Other factors must play a role in accounting for examples like this; we have already mentioned ambiguity avoidance and idiomaticity as two such factors.

Another factor that has been shown to influence constituent ordering is the priming of syntactic structures (Bock 1989, Bock & Loebell 1990). In Bock's (1989) experiments, participants were presented with pictures and asked to describe the depicted event, such as 'The girl is handing a paintbrush to the man on the ladder'. These pictures were preceded by sentences that the participants heard and repeated. The auditory sentences consisted of either a *to*-dative (*A rock star sold some pure cocaine to his manager*) or a *for*-dative (*A rock star got some pure cocaine for his manager*). These sentences were presented in either double-object or prepositional constructions. The results showed that the construction of the priming sentence significantly affected the construction that the

[11] In contrast, when both constituents are pronominal, both orders are acceptable. In the DA corpus there were three items with both theme and goal pronouns: two occurred as double object constructions, one occurred as a prepositional construction.

participants used to describe the picture, regardless of whether it was a *to*-dative or a *for*-dative. While Bock argues that the priming occurred from the syntactic frame, independently of the prepositions themselves, Adele Goldberg (p.c.) suggests that such priming may also be sensitive to semantic similarities between similar syntactic constructions.

In addition to on-line constraints like these, production is obviously influenced by the options provided by the grammar. Researchers have suggested that the shape of a grammar can also be described as a constraint-based system, such as optimality theory (Prince & Smolensky 1993), or head-driven phrase structure grammar (Sag & Wasow 1999). But the type of constraints we have focused on here are performance constraints, those most likely to play a role during on-line planning and production. The relationship between on-line constraints and the constraints that shape a grammar is an issue we leave for other work.

6. Conclusions. The main point of this article is a simple one: both grammatical complexity and discourse status influence constituent order. This result is quite robust, as indicated by the fact that it held up under two different methodologies and two different data sets. In addition, we have argued that the relationship between forms and the underlying processes that determine them is a complex one, with the single formal feature of ordering being influenced by a variety of different psycholinguistic factors. Examination of the factors affecting ordering suggests a processing and production model that employs multiple interacting constraints.

We also considered whether the constraints in question serve the interests of the speaker, those of the hearer, or both. Some of the factors we discussed are likely to affect constituent ordering because they place demands on the production system. The clearest example of this is priming, which could only affect production at a fairly low level. By contrast, other factors may play a role because speakers design their utterances for the benefit of their interlocutors, and they know that certain configurations are easier to interpret or to process than others. This explanation would be the most natural account for the use of constituent ordering as an ambiguity avoidance strategy, which would only benefit the addressee.

The two factors that we investigated, complexity and discourse status, could also play a role in one of two ways. On the one hand, placing given, light material early in the utterance probably facilitates comprehension, so speakers may produce these elements early out of consideration for their interlocutors. On the other hand, as we have argued, these elements are also easier to produce. The results presented here, and in particular the data on disfluencies, suggest that the early production of given, light constituents is influenced by constraints on planning and production.

Appendix: list of all objects

simple objects

key	apple	carrot	peach
melon	tape	balloon	helicopter
eraser	tomato	ruler	lemon
nail	corn	knife	fork
screw	eggplant	banana	sunglasses
strawberry	stapler	gun	slinky
padlock	red pepper	pear	soap
baseball	scissors	plum	celery
disk	glue	razor	sponge

COMPLEX OBJECTS

tops
- green and yellow starred top
- green and yellow plain top
- red and yellow starred top
- red and yellow plain top
- purple and yellow starred top
- purple and yellow plain top

tic tacs
- red and white tic tacs
- red and orange tic tacs
- green and white tic tacs
- orange and white tic tacs

whistles
- blue whistle w/ green string
- blue whistle w/ blue string
- green whistle w/ red string
- green whistle w/ green string
- red whistle w/ blue string
- red whistle w/ red string

crayons
- small green crayon
- big green crayon
- small yellow crayon
- big yellow crayon
- small red crayon
- big red crayon

cups
- silver starred cup
- gold starred cup
- red starred cup
- yellow spotted cup
- red spotted cup
- blue spotted cup

umbrellas
- blue umbrella w/ white handle
- green umbrella w/ white handle
- red umbrella w/ white handle
- blue umbrella w/ black handle
- green umbrella w/ black handle
- red umbrella w/ black handle

spoons
- black spotted spoon
- black starred spoon
- white spotted spoon
- white starred spoon
- yellow spotted spoon
- yellow starred spoon

pencils
- green spotted pencil
- green plain pencil
- yellow spotted pencil
- yellow plain pencil
- pink spotted pencil
- pink plain pencil

cones
- yellow spotted cone
- yellow starred cone
- red spotted cone
- red starred cone
- blue spotted cone
- blue starred cone

spools
- yellow spotted spool
- yellow starred spool
- pink spotted spool
- pink starred spool
- green spotted spool
- green starred spool

cars
- pink striped car
- pink spotted car
- green striped car
- green spotted car
- orange striped car
- orange spotted car

balls
- blue spotted ball
- blue plain ball
- blue striped ball
- orange spotted ball
- orange plain ball
- orange striped ball

REFERENCES

ANDERSON, ANNE H.; ELLEN G. BARD; CATHY SOTILLO; ALISON NEWLANDS; and GWYNNETH DOHERTY-SNEDDON. 1997. Limited visual control of the intelligibility of speech in face-to-face dialogue. Perception and Psychophysics 594.580–92.

ARIEL, MIRA. 1990. Accessing noun-phrase antecedents. London: Routledge.

ARNOLD, JENNIFER E. 1998. Reference form and discourse patterns. Stanford, CA: Stanford University dissertation.

——, and THOMAS WASOW. 1996. The effect of production constraints on particle movement and dative alternation. Poster presented at the CUNY conference on human sentence processing, New York City.

BATES, ELIZABETH, and BRIAN MACWHINNEY. 1989. Functionalism and the competition model. The cross-linguistic study of sentence processing, ed. by Brian MacWhinney and Elizabeth Bates, 3–96. Cambridge: Cambridge University Press.

BEHAGHEL, OTTO. 1909/10. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. Indogermanische Forschungen 25.110–42.

——. 1930. Von deutscher Wortstellung. Zeitschrift für Deutschkunde, Jargang 44 der Zeitschrift für deutschen Unterricht: 81–89.
BEVER, THOMAS G. 1970. The cognitive basis for linguistic structures. Cognition and the development of language, ed. by John R. Hayes, 279–362. New York: Wiley and Sons.
BIRNER, BETTY J. 1998. Recency effects in English inversion. Centering theory in discourse, ed. by Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, 309–23. Oxford: Clarendon Press.
——, and GREGORY WARD. 1998. Information status and noncanonical word order in English. Philadelphia: John Benjamins.
BOCK, J. KATHRYN. 1986. Meaning, sound, and syntax: Lexical priming in sentence production. Journal of Experimental Psychology: Learning, Memory, and Cognition 124.575–86.
——. 1987. An effect of the accessibility of word forms on sentence structures. Journal of Memory and Language 26.119–37.
——. 1989. Closed-class immanence in sentence production. Cognition 31.163–86.
——, and DAVID E. IRWIN. 1980. Syntactic effects of information availability in sentence production. Journal of Verbal Learning and Verbal Behavior 19.467–84.
——, and HELGA LOEBELL. 1990. Framing sentences. Cognition 35.1–39.
——, and RICHARD K. WARREN. 1985. Conceptual accessibility and syntactic structure in sentence formulation. Cognition 21.47–67.
CHAFE, WALLACE L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. Subject and topic, ed. by Charles. N. Li, 25–56. New York: Academic Press.
——. 1994. Discourse, consciousness, and time. Chicago: University of Chicago Press.
CHOMSKY, NOAM. 1970. Deep structure, surface structure, and semantic interpretation. Semantics, ed. by Danny Steinberg and Leon Jakobovits, 183–216. Cambridge: Cambridge University Press.
——. 1975. The logical structure of linguistic theory. Chicago: University of Chicago Press.
CLARK, HERBERT H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior 12.335–59.
——. 1994. Managing problems in speaking. Speech Communication 15.243–50.
——. 1996. Using language. Cambridge: Cambridge University Press.
——, and EVE V. CLARK. 1977. Psychology and language: An introduction to psycholinguistics. New York: Harcourt Brace Jovanovich.
——, and SUSAN E. HAVILAND. 1977. Comprehension and the given-new contrast. Discourse production and comprehension, ed. by Roy O. Freedle, 1–40. Hillsdale, NJ: Lawrence Erlbaum Associates.
——, and CATHERINE R. MARSHALL. 1981. Definite reference and mutual knowledge. Elements of discourse understanding, ed. by Aravind K. Joshi, Bonnie L. Webber, and Ivan A. Sag, 10–63. Cambridge: Cambridge University Press.
——, and THOMAS WASOW. 1998. Repeating words in spontaneous speech. Cognitive Psychology 27.201–42.
ERDMANN, PETER. 1988. On the principle of 'weight' in English. On language, rhetorica phonologica syntactica: A festschrift for Robert P. Stockwell from his friends and colleagues, ed. by Caroline K. Duncan-Rose and Theo Vennemann, 325–39. London: Routledge.
FERREIRA, VICTOR S. 1996. Is it better to give than to donate? Syntactic flexibility in language production. Journal of Memory and Language 35.724–55.
——, and GARY S. DELL. 1998. Syntactic and lexical choices in language production: What we can learn from 'that'. Paper presented at the Eleventh Annual CUNY conference on human sentence processing, Rutgers University.
FIRBAS, JAN. 1966. Non-thematic subjects in contemporary English. Travaux Linguistiques de Prague 2.239–54.
FOX TREE, JEAN E., and HERBERT H. CLARK. 1997. Pronouncing 'the' as 'thee' to signal problems in speaking. Cognition 62.151–67.
FRAZIER, LYNN, and JANET D. FODOR. 1978. The sausage machine: A two-stage parsing model. Cognition 6.291–325.

GIBSON, EDWARD. 1997. Syntactic complexity: Locality of syntactic dependencies. Cambridge, MA: MIT, MS.

GIVÓN, TALMY. 1983. Topic continuity in discourse. Quantitative cross-language studies, TSL #3. Amsterdam: Benjamins.

——. 1988. The pragmatics of word-order: Predictability, importance and attention. Studies in syntactic typology, ed. by Michael Hammond, Edith A. Moravcsik, and Jessica R. Wirth, 243–84. Amsterdam: Benjamins.

GOLDBERG, ADELE E. 1995. Constructions: A construction grammar approach to argument structure. Chicago: University of Chicago Press.

GUNDEL, JEANETTE K. 1988. Universals of topic-comment structure. Studies in syntactic typology, ed. by Michael Hammond, Edith A. Moravcsik, and Jessica R. Wirth, 209–39. Amsterdam: Benjamins.

——; NANCY HEDBERG; and ROGER ZACHARSKI. 1993. Cognitive status and the form of referring expressions. Language 69.274–307.

HALLIDAY, MICHAEL A. K. 1967. Notes on transitivity and theme in English, part 2. Journal of Linguistics 3.199–244.

HAWKINS, JOHN A. 1990. A parsing theory of word order universals. Linguistic Inquiry 21.223–61.

——. 1994. A performance theory of order and constituency. Cambridge: Cambridge University Press.

HOCKETT, CHARLES. 1958. A course in modern linguistics. New York: Macmillan.

HORTON, WILLIAM S., and BOAZ KEYSAR. 1996. When do speakers take into account common ground? Cognition 59.91–117.

KIMBALL, JOHN. 1973. Seven principles of surface structure parsing in natural language. Cognition 2.15–47.

LAMBRECHT, KNUD. 1994. Information structure and sentence form. Cambridge: Cambridge University Press.

LEVELT, WILLEM J. M. 1989. Speaking. Cambridge, MA: MIT Press.

MACDONALD, MARYELLEN C. 1996. Representation and activation in syntactic processing. Attention and performance 16: Information integration in perception and communication, ed. by Toshio Inui and James L. McClelland, 433–56. Cambridge, MA: MIT Press.

——; NEAL J. PEARLMUTTER; and MARK S. SEIDENBERG. 1994. The lexical nature of syntactic ambiguity resolution. Psychological Review 1014.676–703.

NIV, MICHAEL. 1992. Right association revisited. Paper presented at the Thirtieth Annual Meeting of the Association for Computational Linguistics, Newark, DE.

NUNBERG, GEOFFREY; IVAN A. SAG; and THOMAS WASOW. 1994. Idioms. Language 70.491–538.

POSTAL, PAUL. 1974. On raising. Cambridge, MA: MIT Press.

PRINCE, ALAN, and PAUL SMOLENSKY. 1993. Optimality theory: Constraint interaction in generative grammar. (RuCCS Technical Report 2.) New Brunswick, NJ: Rutgers University.

PRINCE, ELLEN. 1981. Towards a taxonomy of given-new information. Radical pragmatics, ed. by P. Cole, 223–56. New York: Academic Press.

——. 1992. The ZPG letter: Subjects, definiteness, and information status. Discourse description: Diverse analyses of a fundraising text, ed. by Sandra Thompson and William Mann, 295–325. Amsterdam: John Benjamins.

QUIRK, RANDOLPH; SIDNEY GREENBAUM; GEOFFREY LEECH; and JAN SVARTVIK. 1972. A grammar of contemporary English. London: Longman.

SAG, IVAN, and THOMAS WASOW. 1999. Syntactic theory: A formal introduction. Stanford, CA: CSLI Publications.

SCHAFER, AMY J., and SHARI R. SPEER. 1999. Intonational disambiguation in sentence production and comprehension. Paper presented at the Twelfth Annual CUNY conference on human sentence processing, New York City.

SIEWIERSKA, ANNA. 1993. Syntactic weight vs. information structure and word order variation in Polish. Journal of Linguistics 29.233–65.

STALLINGS, LYNNE. 1998. Relative weight in the production of heavy-NP shift. Paper presented at the annual meeting of the Linguistic Society of America, New York.

——; 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.

Tomlin, Russell S. 1998. Mapping conceptual representations into linguistic representations: The role of attention in grammar. With language in mind, ed. by Jan Nuyts and Eric Pederson, 162–89 Cambridge: Cambridge University Press.

Trueswell, John C., and Michael K. Tanenhaus. 1994. Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. Perspectives on sentence processing, ed. by Charles Clifton Jr., Lyn Frazier, and Keith Rayner, 155–79. Hillsdale, NJ: Erlbaum.

Vallduvi, Enric. 1990. The information component. Philadelphia, PA: University of Pennsylvania dissertation.

Wasow, Thomas. 1997a. Remarks on grammatical weight. Language Variation and Change 9.81–105.

——. 1997b. End-weight from the speaker's perspective. Journal of Psycholinguistic Research 26.347–61.

Jennifer Arnold
Institute for Research in Cognitive Science
University of Pennsylvania
3401 Walnut St. Suite 400 A
Philadelphia, PA 19102
[jarnold@linc.cis.upenn.edu]

[Received 21 August 1997;
revision received 28 July 1998;
accepted 14 April 1999]