Processing, Prosody, and Optional *to*

Thomas Wasow, Roger Levy, Robin Melnick, Hanzhi Zhu, and Tom Juzek

0. Introduction

Flickinger and Wasow (2013) discuss a previously understudied phenomenon in English that they call the "do-be construction" (DBC). This widely-used construction is characterized by a remarkably rich and interconnected set of constraints. Before enumerating them, we present a few examples[1].

(1) a. the thing that I tried to do was to keep the score close
b. the least we should do is make it as much fun as possible
c. what the CBO does is takes Congress's promises at face value
d. what we have done is taken military action in Bosnia through NATO
e. all he's been doing is going over legal papers

Flickinger and Wasow identify the following as the characteristic properties of DBC:

(2) a. The top verb in the construction is a specificational copula – that is, a form of *be* stipulating identity between the denotations of its subject and its complement.
b. The subject contains a relative clause headed by one of the following seven words: *what, thing, all, best, worst, most,* or *least*.
c. A form of the word *do* occurs within the relative clause.
d. The complement of the copula is a verb phrase (VP).
e. The understood subject of the post-copula VP (PCVP) is the same as the understood subject of *do*.
f. The form of the post-copula verb (PCV) depends on the form of *do* in the subject.

There are many questions one might ask about this construction, including how to analyze it within a particular theory of grammar (Flickinger and Wasow do this for Head-driven Phrase Structure Grammar), what its discourse function is, how it relates to other constructions (e.g., pseudoclefts), how it differs across dialects and registers of English, and what its history is. We will not address any of these here. Rather, we are concerned with what conditions the presence or absence of the infinitival *to* at the beginning of the post-copula verb phrase (PCVP).

As noted in (2f), the form of the post-copula verb (PCV) is constrained. Specifically, there are three possible inflectional forms for the PCV: the same form as *do* in the

[1] Except where otherwise noted, examples in this paper are from the Corpus of Contemporary American English (http://corpus.byu.edu/coca/), or COCA for short. We have truncated many of the examples, keeping only what is needed to make our point. Hence, most of our examples, are presented without initial capitalization or sentence-final punctuation. Invented examples begin with capital letters and end with periods.

subject, base (that is, uninflected) or infinitival (that is, *to* followed by an uninflected verb)[2].

This is illustrated in the contrasts between (1) and (3).

(3) a. The thing that I tried to do was keep/*keeps/*kept/*keeping the score close.
b. The least we should do is to make/*makes/*made/*making it as much fun as possible.
c. What the CBO does is take/to take/*taken/*taking Congress's promises at face value
d. What we have done is take/to take/*took/*taking military action in Bosnia through NATO.
e. All he's been doing is ?go/?to go/*went/*gone over legal papers.

Thus, whenever the PCV can be in base form (without *to*), it could just as well be in infinitival form (with *to*), and vice-versa. To see the apparent interchangeability of these forms, consider the examples in (4), all of which were taken from COCA, but only half of which had *to* in the original.

(4) a. what we're here on earth to do is (to) celebrate humanity
b. what I would do is (to) call upon the press to police yourselves
c. the other thing that it'll do is (to) facilitate getting Chinese troops into Tibet as well
d. the most important thing that Bretton Woods did was (to) create two institutions for international cooperation on monetary international problems
e. all they can do is (to) circumvent themselves
f. all I want to do is (to) go to work

Audiences to whom we have presented these examples do not have clear intuitions about which examples had *to* in the original[3].

This raises the question of what factors lead people to use *to* in the DBC when they do. The bulk of this paper describes a study aimed at answering this question and discussing why the answer is of theoretical interest.

1. **Data Extraction and Annotation**

We conducted a corpus study using COCA, a 450-million word web-based collection, roughly equally divided among speech (radio and television interviews), newspapers, magazines, fiction, and academic writing, dating from 1990 to 2012[4]. COCA is tagged

[2] Flickinger and Wasow claim that if the form of *do* is a present participle (that is, *doing*), then the PCV also has to be a present participle, citing invented examples like the following, which they judge unacceptable:

(i) The thing I'm doing is (to) try to learn from my mistakes.

But the corpus studies we report on here turned up enough real examples similar to (i) to convince us that Flickinger and Wasow were mistaken.

[3] Examples b, d, and f had *to* in the original.

[4] The data in our statistical model were collected in the summer of 2012, when the corpus was somewhat smaller (425 million words) and did not yet have data from 2012.

for part of speech, but not syntactically parsed. It has a user-friendly web interface, which extracts examples based on patterns that may include parts of speech, particular words, disjunction, and wild cards. A small window of context around the matching text can also be extracted.

An earlier pilot study (Wasow, et al, 2012) of optional *to* in the DBC had involved hand-coding 1000 randomly selected examples from the spoken portion of COCA for a variety of factors that we thought might correlate with *to* use. For this paper, our dataset was much larger, including written as well as spoken examples. By using computational tools for extraction, culling, and annotation, we were able not only to obtain considerably more data, but also to consider more factors than in the pilot. These factors are described in Section 2; in the remainder of this section we describe the extraction, culling, and annotation process.

We initially extracted all examples that included some form of the verb *do*, followed by some form of the verb *be*, optionally followed by *to*, and (obligatorily) followed by any verb in base form[5]. The extraction pattern allowed up to two[6] words to intervene between any two of these words – that is, it could be abbreviated as

DO (W)(W) BE (W)(W) (*to* (W)(W)) V[base]

where "DO" means any form of *do*, "W" means one word, and "BE" means any form of *be*. The resulting sample was then parsed with the Stanford parser (Klein and Manning, 2003). Through trial and error, we developed a tgrep2 (Rohde, 2005) pattern to help us cull out examples that were not in fact instances of DBC.

Annotation was done with Perl scripts, some of which made use of the parses. The most obvious need for the parses was in measuring the lengths of constituents, since that required assigning constituent structure. But the parses were also used in identifying such things as the occurrences of *do* and *be* whose forms we thought might influence *to* use. The annotations provided by the scripts were subsequently used to automatically code the data for the factors we considered for use in modeling. In some cases, the annotations could simply be used as codings (for example, the form of *do* in the subject and whether the example was written or spoken), but in others some additional computation was required -- e.g., the measure of subject length was computed by subtracting the position of the subject's head noun in the sentence (that is, its distance from the start of the sentence) from the position of *do* in the sentence. These

[5] COCA has two distinct tags verb.BASE and verb.INF for uninflected non-finite verbs. We have not been able to discern a consistent basis for this distinction, although verb.INF seems to appear after *to* at a considerably higher rate than verb.BASE. In all of our searches, we used the disjunction of these two tags to search for what we call base forms of verbs. For the purposes of this paper, we treated the two COCA tags as interchangeable. That is, when we say a verb's form is base, we mean it is uninflected and not preceded by *to*; and when we say a verb is infinitival, we mean it is preceded by *to*.

[6] The limitation of at most two intervening words was required for computational reasons.

computations were carried out by an R script, which also renamed some of the annotations and removed unused fields.

Some codings (e.g., ones that didn't give one of our seven nouns as the head noun of the subject or that gave the number of words between *do* and *be* as more than 2) triggered hand-checks of particular examples, and additional random hand-checks were performed. Altogether, we hand-checked hundreds of examples and discarded examples that were not the type of DBC sentences we were investigating. Our final dataset contained 10116 examples, but 143 of them had uncoded values for some variable used in our analysis. Furthermore, only one example involved the *were* form of the copula. We dropped these 144 examples before statistical analysis, so that our analyses involved 9972 examples. In a random check of over 100 examples, all were examples of the DBC with base or infinitival post-copula verbs.

We used the same pipeline to extract and annotate DBC examples from the Fisher corpus (Cieri, et al 2004). Fisher consists of telephone conversations on designated topics; it is far smaller (about 22 million words) than COCA. Analysis of the Fisher dataset was qualitatively consistent with the COCA model we report on below, but the number of examples extracted (861) was too small to show reliable effects for many of the significant factors in our COCA data. Consequently, we provide a detailed accounting only of the COCA study.

2. **Factors in our Analysis**

Based on earlier work on optional *that* in both relative clauses and complement clauses (see Jaeger, 2010 and Wasow, et al, 2011, *inter alia*), we expected that similar factors might influence the presence or absence of *to*. In particular, we expected factors that contribute to the processing difficulty of a DBC sentence would increase the probability of *to* use. These factors include long and/or syntactically complex phrases within the sentence. They also include the use of relatively infrequent words or word forms.

Why should processing difficulty encourage *to* use? The obvious answer is that the extra little word takes time, giving the speaker an extra fraction of a second for planning the remainder of the utterance and lexical retrieval. The extra time is also useful for the listener, providing more time for parsing and lexical retrieval. The work on *that* suggests that these effects show up in writing as well as in speech, even though our hypotheses about why they occur are based on the temporal pressures on speakers and listeners. This could be due either to habits of speech being preserved in writing, or to similar temporal pressures on readers. We will not attempt to resolve this question here.

We coded measures of phrasal complexity and word frequency based on the parts of the utterance most closely connected with the site of optional *to* and thus most likely a priori to influence speaker choice, where by "connected" we mean parts of the utterance that are components of the DBC (see (2) above) and/or are close to optional *to* in terms of linear ordering. For phrasal complexity, this led us to code the amount of material in (i) the subject NP between the head noun and *do*, (ii) between *do* and *be*, (iii) between *be*

and the PCV, and (iv) in the post-copula verb phrase (PCVP). We expected that in all cases more material would lead to greater utterance complexity and thus greater preference for *to*. Both length and complexity can, of course, be measured in multiple ways. For length we used number of words, though number of syllables might have been as good or better, as might duration (for speech) or number of characters (for writing). There is a substantial body of literature (see, e.g. Hawkins 1994 and Wasow 2002) that has found number of words to be a good proxy for more sophisticated measures of complexity. Complexity measures tend to depend on the parse assigned, and the ones we had were not very reliable. Moreover, since complexity is highly correlated with length, the only complexity measure we looked at was number of verbs in a phrase. This turned out to be highly collinear with length and a less reliable predictor, so we ended up relying only on number of words for our length/complexity measures.

We also examined the effects of wordform frequencies[7] for critical components of the DBC: the head of the subject NP, the form of *do*, the form of the specificational copula *be*, and the post-copula verb (PCV). For the first three of these, only a small number of wordforms are possible, so in our analysis we directly modeled the *to*-use preference associated with each wordform and performed exploratory visualizations of the relationship between preference and in-construction frequency of the wordform (Section 3.3).

In contrast, there are many different PCVs; furthermore, the PCV is distinctive among DBC components in that there is strong reason from a mathematically precise theory for predicting that its frequency will affect *to*-use preference. Namely, the theory of Uniform Information Density (UID; Levy & Jaeger, 2007; Jaeger; 2010) posits that communicative efficiency is optimized if information is transmitted at a uniform rate, and that speakers take advantage of the grammatical opportunities afforded them to smooth this information rate out. The notion of "information" here is based on information theory (Shannon, 1948), and is measured as log of inverse probability (equivalently, negative log-probability) or *surprisal*,. It follows that optional function words like *that* and *to* are more likely to be inserted in environments where, without them, there might be an information peak[8]. To understand how this applies to the DBC, we can use reasoning directly analogous to that developed by Levy and Jaeger (2007) for *that*-use in relative clauses; the key is that the PCV is often the first point in the utterance where it becomes clear that the utterance must involve a DBC. Consider the variant of Example (3a) without *to*:

(5) what we're here on earth to do is *celebrate* humanity

Before the PCV *celebrate*, there are alternative ways that the utterance could continue

[7] We used frequencies of these forms in our sample, rather than in the whole of COCA.

[8] To test whether people employ this Uniform Information Density (UID) strategy in actual usage using corpus studies has required computing information at critical points in utterances on the basis of very local information, usually immediately preceding n-grams for some very small n.

that do not involve the DBC:

(6) (a) what we're here on earth to do is a complete mystery
(b) what we're here on earth to do is unique
(c) what we're here on earth to do is not what you think we're here to do

Therefore, in the *to*-free variant, the PCV conveys two distinct pieces of information about the structure and content of the utterance: (i) the fact of the DBC, and (ii) the identity of the PCV of the DBC. These can be measured information-theoretically as follows:

$$\log \frac{1}{P(\text{PCV,DBC}|\text{Context})} = \overbrace{\log \frac{1}{P(\text{DBC}|\text{Context})}}^{\text{(i)}} + \overbrace{\log \frac{1}{P(\text{PCV}|\text{DBC, Context})}}^{\text{(ii)}}$$

The use of *to* separates out these two pieces of information: *to* conveys (i), whereas after *to* the PCV conveys only (ii). Therefore, the optimal distribution of optional *to* from an information-density perspective would to use it when (i), (ii) or both are large. With respect to PCV information content, this line of reasoning predicts that *to* use in the DBC will be higher when the PCV is less predictable, and (ii) is thus large. In principle (ii) should be measured with respect to the complete context; but a reasonable and convenient first simplification is to assume that P(PCV|DBC,Context) ≈P(PCV|DBC)—namely, that in-construction PCV frequency allows us to approximate the information content of (ii).

We also expected that priming could increase the probability of *to*, so we expected that, when *do* in the subject was in infinitival form (that is, preceded by *to*) the rate of *to* before the PCV would be increased.

An expectation that was not derivative from the work on optional *that* was that some phonological factors might influence the use of *to*. This idea was suggested to us by Arto Anttila, who has shown the influence of prosody on other syntactic alternations in English (e.g. Anttila, et al 2010). He also brought to our attention a book published over a century ago entitled *Rhythm in English Prose* (van Draat, 1910) with a chapter entitled "The infinitive *with* and *without* preceding *to*"[9], which argued that *to* could have a prosodic function. Based on Anttila's suggestion, we considered whether *to*, which is virtually always unstressed, might sometimes serve to prevent two stressed syllables from appearing adjacent to one another, a situation known to be disfavored and referred to as "stress clash" (see Liberman and Prince, 1977). To understand how this might influence speaker choice regarding *to* production, consider the following two examples from the spoken section of COCA, with the presumably stressed syllables in bold (see the next paragraph regarding stress status of the copula *is*):

[9] Interestingly, all of the cases discussed in van Draat's chapter, except the complement of *help* now strike us as categorically either requiring or prohibiting *to*.

(7) And **one** of the **best ways** to **do** it **is (**to) **break bread** with **them**.
(8) **All** I can **do is** (to) con**ti**nue to be**have** in a **way** that **earns** your **trust**.

In both cases, the inclusion or omission of *to* has no bearing on the grammaticality of the sentence. However, speakers' *to*-use decisions could affect the prosodic optimality of the utterances. In (7), omitting *to* would cause a stress *clash* – a sequence of two consecutive stressed syllables, ***is*** and ***read*** – that would be avoided by the use of *to*. In (8), on the other hand, including *to* would cause a stress *lapse* – a sequence of two consecutive unstressed syllables, *to* and *con-*, which would be avoided by the omission of *to*. If speakers are sensitive to this potential prosodic function of *to*, they should tend to include *to* when its omission would cause stress clash, and omit *to* when its omission would cause stress lapse. (In fact, *to* was used in (7) and omitted in (8) in COCA.) Note that in cases where nothing (other than *to*) intervenes between the copula and the PCV, the predictions of clash avoidance and lapse avoidance are identical: a PCV with initial stress should favor *to* use more than a PCV with non-initial stress. Since this covers a large majority of our examples, we conflated clash avoidance and lapse avoidance into one factor, which we refer to as clash avoidance, or simply (potential) stress clash.

We determined stress clash by annotating (i) the PCV for whether it had initial stress and (ii) the word immediately preceding the PCV (or the word immediately preceding *to*, if *to* is used) for whether it had final stress. Since the word immediately preceding the PCV is normally a copula – usually *is* – our ability to investigate potential effects of stress clash hinges on whether the copula is stressed. Is it? If so, it is not clearly audible. On the other hand, the fact that *is* in the DBC is never contracted (and sounds quite unacceptable when contracted, e.g., **All you need to do's pay attention*) suggests that it does carry some stress. We thus considered the copula as stressed in our dataset

In cases where something (other than *to*) intervenes between the copula and the PCV, these arguments based on prosody depend on the stress pattern of the intervening material. The situation is somewhat complicated by the fact that, when *to* appears, intervening material can appear before and/or after *to*. Hence, when there is intervening material but no *to*, there may be multiple locations where insertion of *to* would be grammatical, and the effect on prosody would be different in each one. To avoid this complication, we made the simplifying assumption that, in cases without *to*, the alternative we were comparing the actual sentence to was one with *to* immediately preceding the PCV. Because the vast majority of examples do not have material intervening between the copula and the PCV, this simplification is unlikely to have materially affected our results.

In addition to prosody, we thought segmental phonology might, conceivably, influence the use of *to*. Our reasoning was that, if the initial segment of the PCV is too similar phonologically to the final segment of the preceding word, the word boundary might be obscured. We conjectured that such a situation might favor *to* use. Consequently, we coded for the initial segment of the PCV and for the final segment of the preceding word.

The final factor we thought might affect *to* use is whether the sentence in question is

spoken or written. Our sample contained roughly equal numbers of examples from speech (4865) and writing (5251), and we included this factor as one we considered. We did not actually know what to expect in terms of this factor's effects. On the one hand, the DBC occurs at a much higher rate in speech than in writing (recall that COCA is 80% written), and written language tends to employ more complex structures and longer sentences than speech. These factors would lead to the expectation of a higher rate of *to* use in writing than in speech. On the other hand, if one of the reasons for using *to* is to buy time in production, then it should appear more frequently in speech. It turns out that speech favors *to* use: *to* occurs in 38% of our spoken examples, compared to 29% of written examples.

The following is a list of the factors we used in our analyses:

- Head noun of the subject. We had four values for this: ALL, THING, WHAT, and SUPER (for "superlative"), where the last category includes the relatively rare head nouns *best, worst, most,* and *least*, plus the handful of examples with something else heading the subject.
- Subject length. This was measured in words, from the head noun to *do*.
- Form of *do*. We considered seven forms: base, infinitive (with *to*), present tense non-third person *do, does*, *did, done,* and *doing*.
- Form of the copula. The vast majority of the examples have *is*, but *was* also occurs with some frequency, and there are some examples with *are*.
- Number of intervening words between *do* and *be*. In the COCA data, this could be zero, one, or two, with most cases being zero.
- Number of intervening words between *be* and the PCV. Again, in the COCA data, this could be zero, one, or two, with most cases being zero.
- PCVP length. This was measured in words, including the PCV (but excluding *to* when present). It relied on the parse tree to find the end of the PCVP.
- Frequency of the PCV in our sample. As is standard in corpus studies, we used the log of the frequency.
- Stress clash. This would occur (without *to*) if the PCV had initial stress and the preceding word had final stress. We treated the copula has having final stress.
- Segmental phonology. We classified the initial segment of the PCV and the final segment of the preceding word (not counting *to*, when present) into one of four categories: vowels, sibilants, sonorants, and other. We then coded each example for whether the two segments in question were of the same or different categories.
- Speech vs. writing.

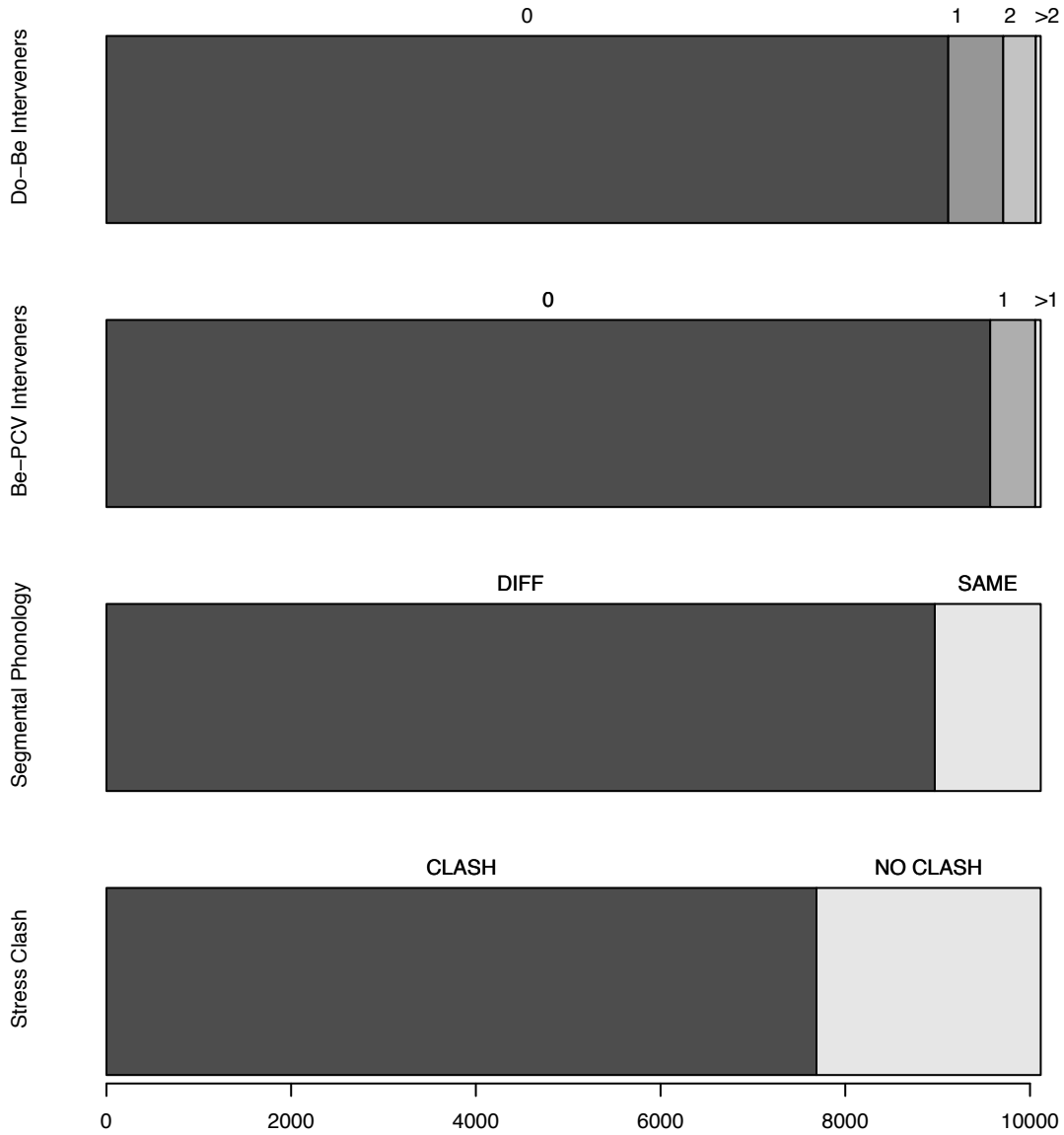

Figure 1: Univariate statistics for *do-be* interveners, *be-PCV* interveners, segmental phonology of initial PCV segment and final segment of preceding word, and potential stress clash

Figure 1 shows univariate statistics for four of these factors: number of *do-be* and *be-*PCB interveners, segmental phonology, and stress clash. Univariate statistics for other factors can be found in Sections **3.3** and **3.4**.

3. Model of the Data

3.1 Mixed logit models

To analyze the effects of these various factors in our data, we use *mixed-effects* (sometimes also called *hierarchical* or multi-*level*) *logistic regression* analysis (or *mixed logit* analysis for short; Pinheiro & Bates, 2000; Bresnan et al., 2007; Baayen et al., 2008; Jaeger, 2008). Mixed logit analysis uses data to infer the dependence of a single,

dichotomous *response variable*—in our case, whether the optional word *to* is used in a given utterance—on one or more *predictors*, allowing for the possibility that different factors may have overlapping and even interacting influences on the response variable. In particular, mixed logit analysis follows the assumption of basic logistic regression (Cedergren & Sankoff, 1974; Agresti, 2002) that all effects of and interactions among predictors can be expressed in terms of additive effects on the *log-odds* of the outcome of the dependent variable; these effects are the *regression coefficients* and inferred from data. For example, consider two hypothetical utterances differing only in the head noun of the subject:

(9) (a) All she did was (to) stare and smile
(b) What he did was (to) stare and smile

If the difference in the regression coefficients associated with *what* and *all* were, for example, log(4)=1.39, then the difference in the log-odds of *to*-use between the examples would also be 1.39. Additive effects on log-odds can equivalently be expressed as multiplicative effects on odds, so the ratio of the odds of *to*-use in the two examples would be $e^{\log(4)}=4$: if the odds of *to*-use were 1:2 for (i) (33% chance of using *to*), then the odds for (ii) would be 2:1 (67% chance); if the odds were 1:1 for (i), then the odds for (ii) would be 4:1 (80% chance), and so forth. We code our dependent variable and predictors such that positive regression coefficients indicate favoring *to*-use, whereas negative coefficients indicate favoring *to*-omission.

Mixed logit analysis extends this picture by adding to the "fixed effects" of ordinary logistic regression a set of "random effects": idiosyncratic departures from the "overall" population norm in baseline behavior and sensitivity to predictor variables that vary across meaningfully clustered subsets. In our case, it is the PCV that makes mixed logit analysis essential. Since the presence or absence of *to* is a feature of an utterance highly local (in both linear and structural terms) to the PCV, it is quite plausible that different PCVs might possess idiosyncratic preferences regarding baseline level of *to* use due to historical accident and/or systematic pressures that we have not measured and included in our model. Furthermore, PCVs have a nearly Zipfian distribution in our dataset (2), so that some PCVs are attested in dozens or even hundreds of utterances. If PCVs do in fact have idiosyncratic *to*-use preferences—for example, if *get*, which occurs in over 900 observations, idiosyncratically prefers *to* more strongly than *make*, which occurs in nearly 500—not including such preferences in our model will interfere with the inferences we draw regarding the effects of other predictors. Finally, note that several theoretically critical predictors in our model—including in-construction PCV frequency, potential stress clash, and segmental phonology—are nearly completely determined by which PCV occurs in the construction.[10] By including a by-PCV "random intercept" in our model, we avoid the "language as fixed effect fallacy" (Clark, 1973; Barr et al., 2013) and ensure that, if we conclude that these predictors reliably affect *to*-use, it is above and beyond any apparent patterns that might emerge due to idiosyncratic PCV-specific

[10] We say *nearly* because in the infrequent cases when material such as adverbs intervenes between the copula and the PCV, stress-clash and segmental phonology predictors are determined by that material, not by the PCV.

preferences alone. For the same reasons, we include a by-PCV "random slope" for the effect of corpus type in our model, since there could be PCV-specific differences between speech and writing in *to*-use preferences. The complete formal specification of our mixed logit analysis is as follows: the probability of *to*-use in a given utterance with fixed-effects predictors denoted by $x_1, \ldots, x_n$ is

$$P(to) = \frac{e^{\eta}}{1 + e^{\eta}}$$

where the *linear predictor* η is

$$\eta = \alpha + \beta_1 x_1 + \cdots + \beta_n x_n + a_{PCV} + b_{PCV} CorpusType$$

and a_{PCV} and b_{PCV} are jointly normally distributed PCV-specific regression coefficients. We fit our model using version 0.999999-2 of the lme4 package of R (Bates et al., 2013), which estimates mixed logit models by maximizing Laplace-approximated data likelihood.

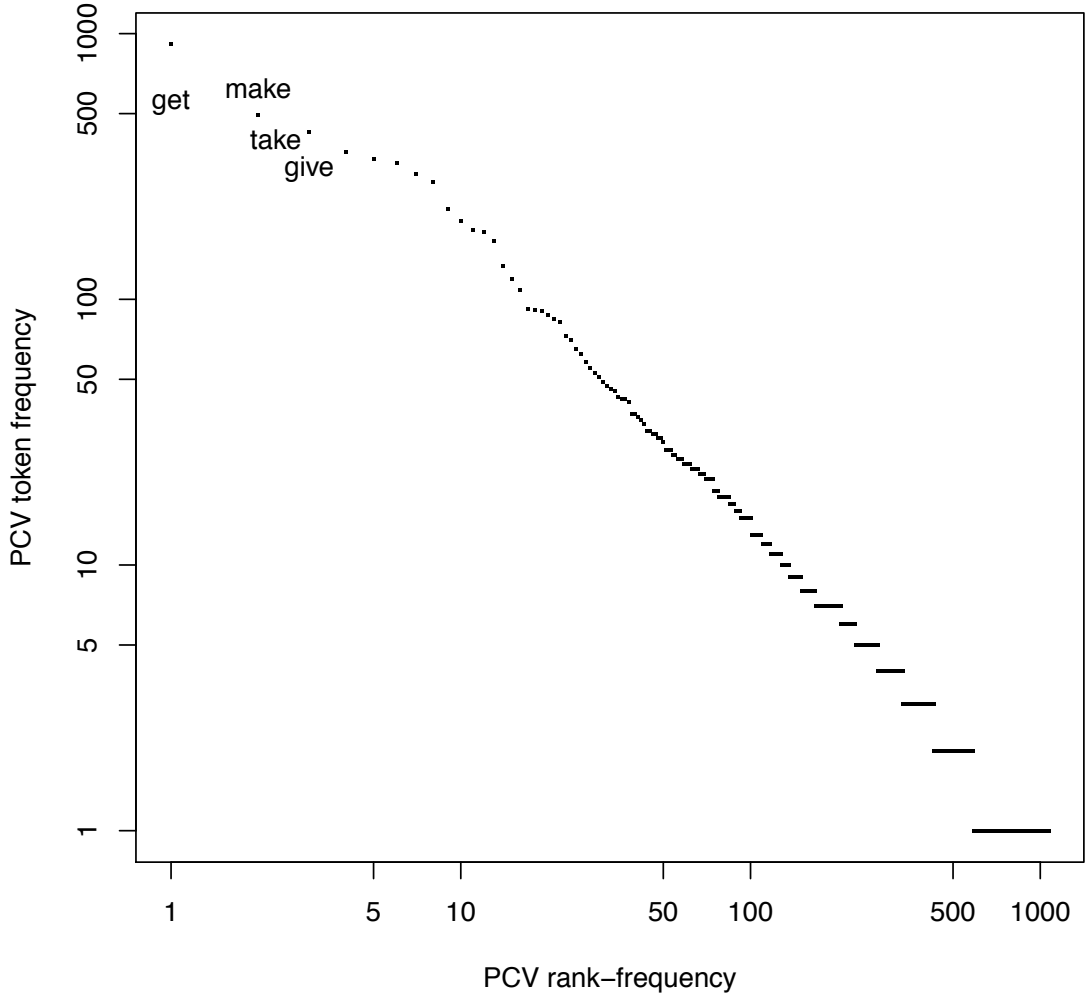

Figure 2: the near-Zipfian distribution of post-copular verbs in the DoBe construction

3.2 Base model results

Our fitted regression model is given in Table 1. Several of the factors in our model are non-numeric, specifically: head noun of the subject, form of *do*, form of the copula, stress clash, and speech vs. writing. For Table 1, we employed what is known as "treatment coding" (Chambers & Hastie, 1993) for these factors: one value of the factor is arbitrarily selected as the baseline, and each of the other possible values appears as a separate predictor in the regression, with the coefficient representing the difference in effect on *to*-preference between the value in question and the baseline value for the factor. These baseline values are ALL for subject NP head, base *do* for *do*-type, *is* for *be*-

form, SPOKEN for corpus type, and CLASH for stress. For the continuous predictors, in our initial model fit we assume simple linear effects on the linear predictor, but explore possible nonlinear effects on *to*-use later in this section. Positive values in the first column of Table 1 indicate that the predictor correlates positively with *to* use, all other predictors being held constant; negative values indicate a negative correlation with *to* use. The absolute value of the regression parameter estimate in the first column indicates the strength of the predictor's effect, and the final column gives a measure of statistical significance based on the Wald *z* statistic.[11]

```
Random effects:
 Groups Name               Variance Std.Dev. Corr
 PCV    (Intercept)        0.990401 0.99519
        corpus.typeWRITTEN 0.028768 0.16961  0.565
Number of obs: 9972, groups: PCV, 1060

Fixed effects:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)          -2.287661   0.166618 -13.730  < 2e-16 ***
Subj.NP.headSUPER     2.579917   0.108405  23.799  < 2e-16 ***
Subj.NP.headTHING     1.871227   0.086696  21.584  < 2e-16 ***
Subj.NP.headWHAT      1.421681   0.082513  17.230  < 2e-16 ***
Subj.length.posthead  0.081163   0.020336   3.991 6.58e-05 ***
do.be.interveners     0.430031   0.048972   8.781  < 2e-16 ***
be.PCV.interveners   -0.021152   0.095580  -0.221    0.825
PCVP.length           0.034689   0.004176   8.307  < 2e-16 ***
do.typedid            0.547730   0.126449   4.332 1.48e-05 ***
do.typedoes           0.219956   0.176829   1.244    0.214
do.typedoing          3.158128   0.595429   5.304 1.13e-07 ***
do.typedone           1.661955   0.131475  12.641  < 2e-16 ***
do.typefinite do      0.172856   0.135887   1.272    0.203
do.typeinf do         0.939312   0.081400  11.539  < 2e-16 ***
be.formare            0.303157   0.547914   0.553    0.580
be.formwas            0.836636   0.075825  11.034  < 2e-16 ***
corpus.typeWRITTEN   -0.275712   0.060109  -4.587 4.50e-06 ***
seg.phonSAME         -0.038220   0.122963  -0.311    0.756
stressNO CLASH       -1.012925   0.106449  -9.516  < 2e-16 ***
```

Table 1: overall model fit

We first summarize those results that can readily be understood from Table 1, and later proceed to explain results that require further visualization. To begin with, the random-effects part of our fitted model assigns considerable idiosyncratic variability across PCVs in *to*-preference not captured by other predictors in our model: the random by-PCV

[11] Each major predictor statistically significant in Table 1 is also significant by a likelihood-ratio test in which the null hypothesis includes a random by-PCV slope for the predictor (results not shown).

intercept has standard deviation 1.00 (units on the logit scale).[12] However, the idiosyncratic difference in *to*-preference of any given PCV in written versus spoken usage is very small (standard deviation 0.17): PCV-specific *to*-use preferences are consistent across genre.

Moving on to fixed effects, we see several critical pieces of evidence supporting our general predictions. Our general prediction from the perspective of utterance planning was that factors increasing memory load and planning difficulty should also increase the rate of *to*-use. In Table 1 we see this prediction confirmed in the positive parameter estimates for the effects of the post-head length of the subject NP, the number of words intervening between *do* and *be*, and the length of the PCVP. All of these estimates differ significantly from zero at $p<0.005$ or more highly significant. We also see confirmation of the more specific prediction of Uniform Information Density (UID): the higher the conditional log-probability of the PCV given the preceding context (here, crudely approximated by conditioning on the fact of being in the DBC), the less likely *to* is to be used. For the continuous predictors and for speech vs. writing, we can see from the table that the correlations are all in the direction predicted. Thus UID and the more general hypothesis that difficulty in utterance planning favors *to*-use receives broad empirical support. The exception is that the presence and number of interveners between *be* and the PCV has no effect on *to*-use preference in this model—but see Section 3.5 for further discussion of this predictor.

We also explored two predictions regarding the effects of phonological predictors on *to*-use. The predicted effect of segmental phonology—namely, that the first segment of the PCV and the final segment of the immediately preceding word being of the same type would promote *to*-use—was not borne out. However, the predicted effect of prosody—that when the first syllable of the PCV and the final syllable of the immediately preceding word are both stressed, *to*-use would be favored to eliminate stress clash—was strongly confirmed. This can be seen in Table 1 from the fact that the parameter estimate associated with NO CLASH is negative (with respect to the baseline level of CLASH).

3.3 Categorical predictors in greater detail

We now examine in greater detail the effects of categorical predictors with more than two levels: subject NP head, *do* type, and *be* form. Although Table 1 contains all the information necessary to reconstruct the effect of each of these predictors, it is not the easiest format in which to visualize these effects, in particular because the degree of confidence in the size and direction effect of the "baseline" level of each predictor is not visible. For this reason, the next series of figures provides visualizations of these effects based on "sum" or "deviation" coding of predictor levels, where the effect of each predictor level is estimated subject to the constraint that the sum is zero. This

[12] To perhaps give a better sense of effect sizes seen in our regression model, a difference of 1 unit on the logit scale is equivalent to the difference between *to*-use probabilities of, for example, 0.02 and 0.05, between 0.05 and 0.12, between 0.12 and 0.27, or between 0.27 and 0.5.

representation also allows us to explore our general hypothesis that low-frequency material favors *to*-use due to difficulty in utterance planning and production. Because the subject NP head, *do*, and the copula are all critical components of the *do-be* construction, it is likely that they would have similar influences as the PCV: the more frequently the particular variant of each component occurs in the construction, the more it should favor *to*-omission. We visualize the extent to which each predictor's effects conform to this hypothesis by passing weighted best-fit lines through the estimated effects in each plot (Figure 3 through Figure 5).[13] Since each of these three components has only a small number of variants, our results regarding relationship of variant frequency against *to*-use preference are necessarily exploratory but, as will be seen momentarily, are provocatively consistent with our general theoretical predictions.

Figure 3 shows the effects of different subject NP heads on *to*-use preference. As predicted by our general hypothesis, we see a general trend for more frequent subject NP heads to dispref*er* *to*-use more strongly. However, the dispreference of the head *all* for *to* is far stronger than would otherwise be expected from this tendency. We have no explanation at present for this exception.

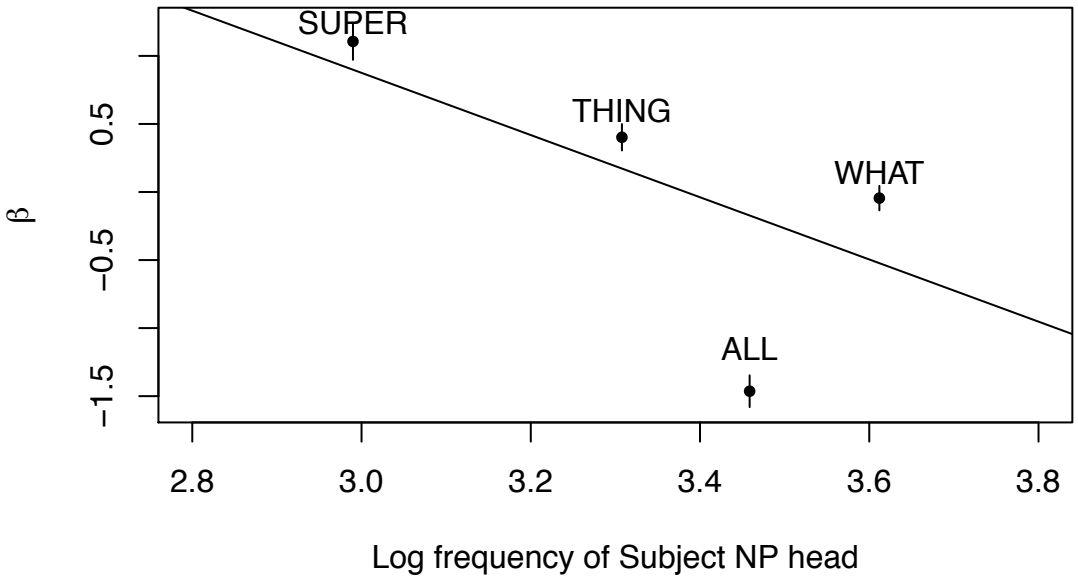

Figure 3: The effects of different subject NP heads on *to*-use preference. Error bars are standard errors of the regression parameter estimate.

The form of *do* is a particularly interesting factor, as shown in Figure 4. Our prediction

[13] The weights for the best-fit line are the inverses of the squared standard errors of each parameter estimate.

that more frequent forms[14] of *do* would have lower rates of *to* use holds up well, except for one severe exception: when *do* is infinitival (that is, *to do*), the use of *to* with the PCV is much higher than would be predicted on grounds of frequency. This exception, however, is consistent with another psycholinguistically motivated prediction we made: that the *to* in the infinitival *do* primes the later use of *to*. Thus, Figure 4 nicely matches our predictions

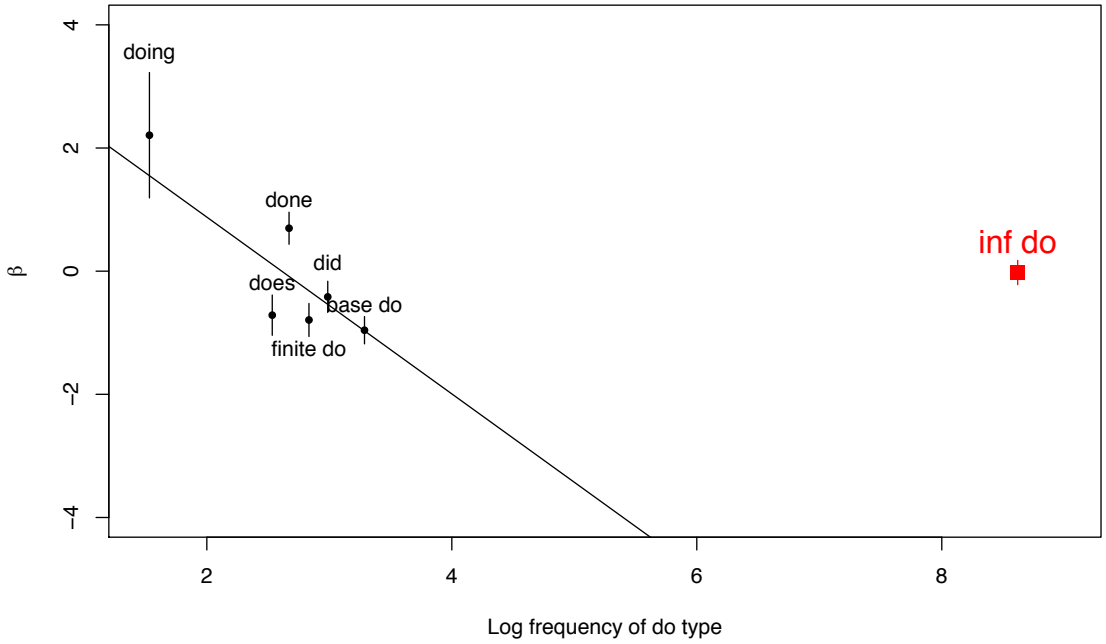

Figure 4: The effects of different *do* types on *to*-use preference. Error bars are standard errors of the regression parameter estimate.

Figure 5 shows the effects of different forms of the copula on *to*-use preferences. Although there are only three distinct forms in our dataset,[15] and one of them, *are*, is so infrequent that our model has little confidence in its precise effect, the general trend, driven by relative preferences for *is* and *was*, is for more frequent copula forms to be associated with less *to*-use, once again consistent with our general hypothesis.

[14] Frequency is measured as the number of occurrences of the form in question as the obligatory *do* of the DBC in our dataset.

[15] Note that we discarded the one instance of a *were* copula since one instance is insufficient data to estimate that form's effect.

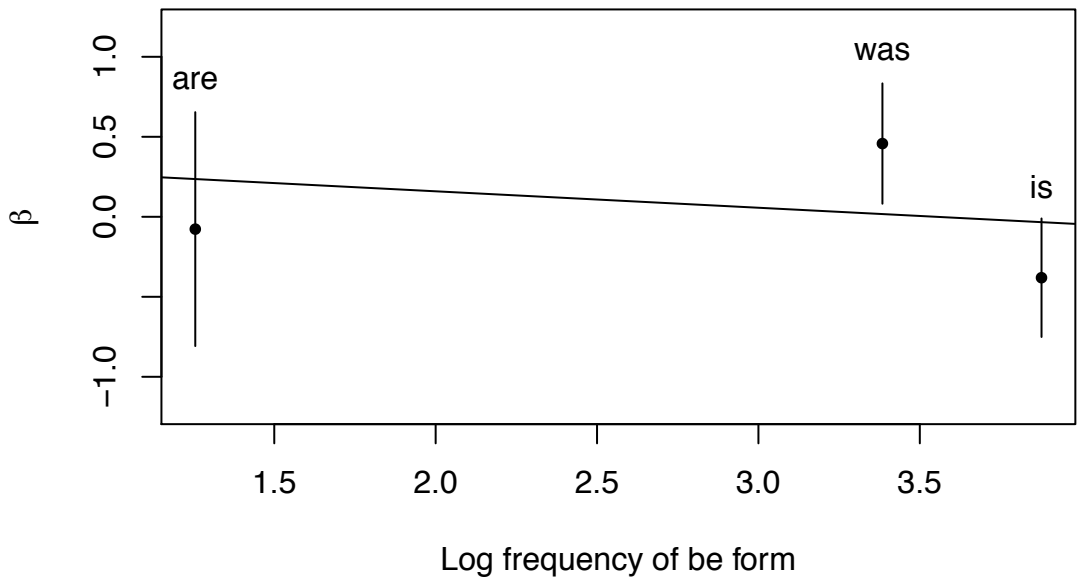

Figure 5: The effect of *be* form on *to*-use preference. Error bars are standard errors of the regression parameter estimate.

In sum, in all four "critical" components of the DBC – the subject NP head, the form of subject NP-internal *do*, the form of the main-clause copula, and the choice of PCV – we find the same pattern emerging: the lower the in-construction frequency of the variant of a component, the more strongly the variant favors *to*-use. One clear exception to this generalization, that infinitival *do* does not disfavor *to* despite its being far and away the most common *do* form, has an independent explanation, namely that it induces repetition priming of *to*-use in the main clause. Hence we see broad support from construction component frequencies for our hypothesis that utterance complexity favors *to*-use.

3.4 Continuous predictors in greater detail

We now move on to a more detailed investigation of the effects of the continuous predictors for which we found significant effects on *to*-use preference in the base model of Table 1. Our depth of understanding of these effects is limited by the assumption built into this base model that the effects of these predictors are linear in log-odds space. For each of these predictors, we explored their effects on *to*-use in more depth by relaxing this assumption: we removed the predictor from the basic model of Table 1 and putting in its place a richer version of the predictor using restricted cubic splines (Green & Silverman, 1994), which allows the model to learn nonlinear effects of the predictor on the log-odds of *to*-use. Figure 6 through Figure 8 depict these effects, together with 95% confidence intervals, controlling for the effects of other predictors; as with Table 1, more positive values indicate stronger preference for *to*-use. At the bottom of each figure is a summary of the data distribution for the predictor in question: for discrete predictors

(subject and PCVP length) a histogram of counts among the 9972 total in the model, and for the continuous predictor of in-construction PCV frequency a kernel density estimate.

Figure 6 shows the results for the post-head length of the subject NP. This figure reveals a regularity invisible in the base model of Table 1: that the general pattern of longer subject NPs favoring *to*-use is reversed for very short subject NPs with four or fewer post-head words. The reason for this reversal is currently unclear to us. One speculative suggestion is as follows: in many utterances with 3 or 4 post-head subject NP words, the only material beyond the minimum (which is two words: a single-word subject of the relative clause, and *do*) is auxiliary and/or modal verbs preceding *do*, which may not add appreciably to utterance complexity (Warren & Gibson, 2002). This speculation would require further research to investigate seriously, however.

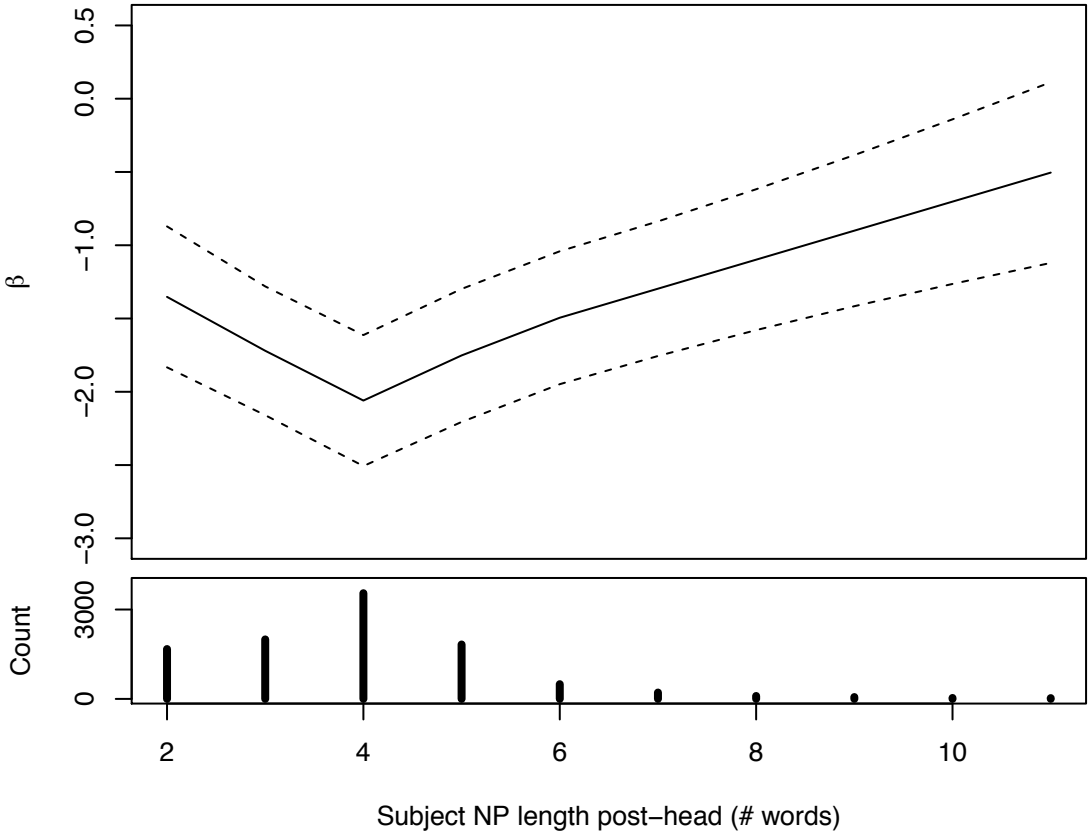

Figure 6: effect of the number of post-head words in the subject NP on *to*-use preference

The effect of the length of the PCVP is illustrated in Figure 7. We see a near-linear effect of PCVP length on *to*-use preference throughout its range; the linearity assumption of the base model in Table 1 was in fact reasonable.

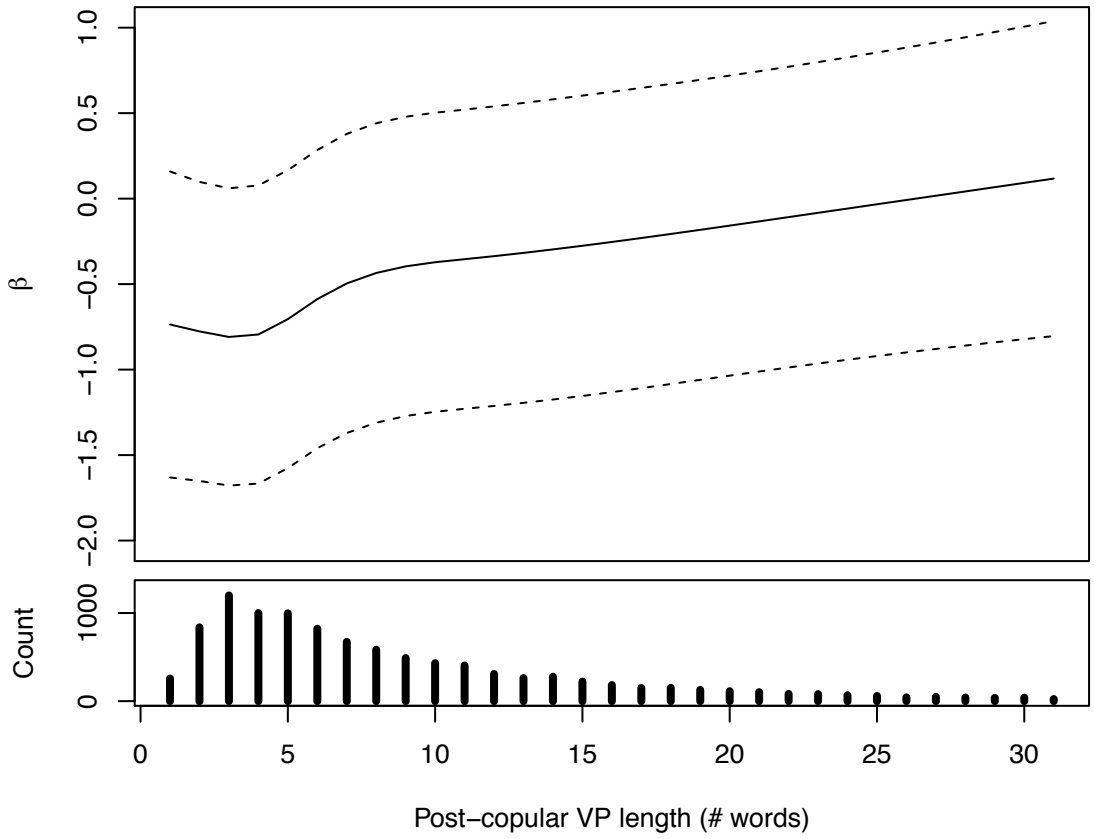

Figure 7: Effect of PCVP length on *to*-use preference

Figure 8 shows the effect of in-construction PCV log-frequency. As with PCVP length, we see a near-linear effect throughout the range of PCV frequency (though model confidence in effect shape drops off for the sparse, highest-frequency PCV range), validating the linearity assumption of the base model in Table 1, which ultimately derived from the theory of Uniform Information Density.

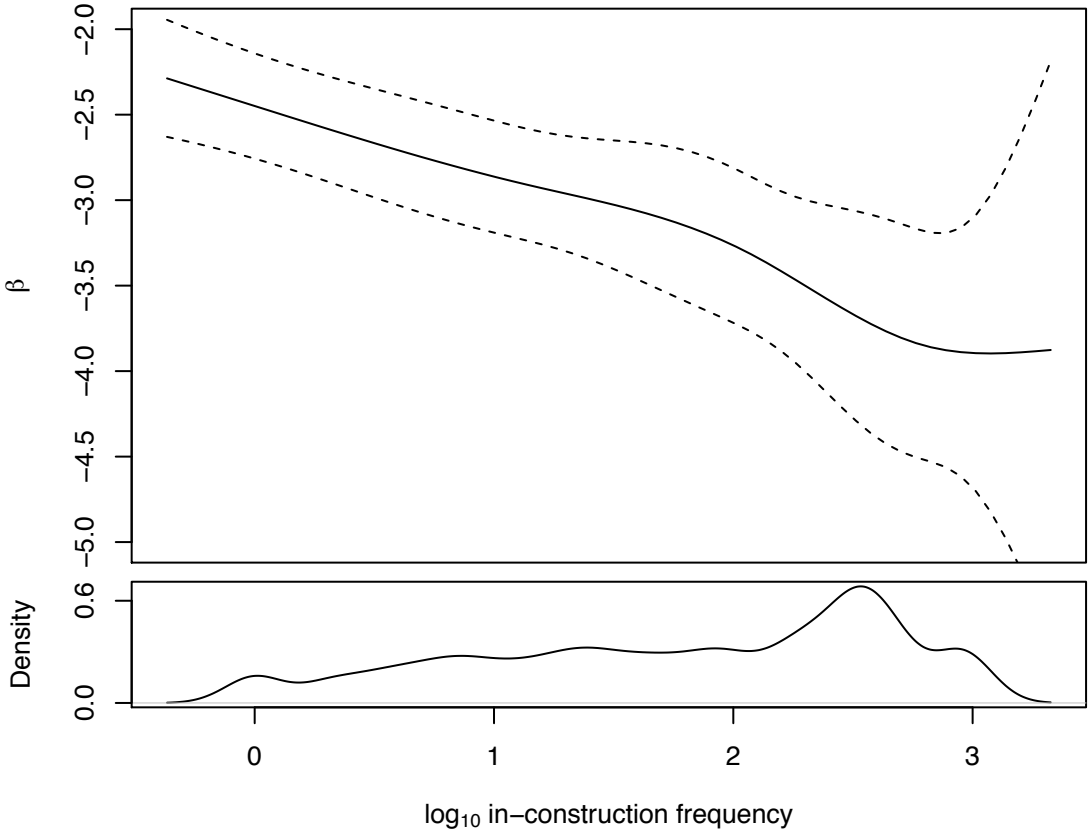

Figure 8: the effect of in-construction PCV frequency on *to*-use preference

In sum, more in-depth spline-based analyses of our continuous predictors largely validate the linearity assumption implicit in the base model of Table 1. The one exception is that there is a reversal of the subject NP post-head length effect for the shortest subject NPs, a pattern whose source we have speculated on but would require further research to understand more fully.

3.5 Interactions with corpus type

Written language shows at least one difference from speech in its lower overall *to*-use rates; moreover, it is not subject to the same real-time production pressures as speech and does not normally indicate what should be stressed (other than occasional marking of contrastive stress through devices like boldfacing). Thus, it is natural to ask whether the effects of complexity and stress clash avoidance in our model may differ between speech and writing. To answer this question, we tested for significant interactions between corpus type and each of the other predictors in our model, using likelihood-ratio test model comparison in each case between the base model of Table 1 and a minimally enriched model in which only an interaction between corpus type and the predictor in

question was added. Our predictors fell into three categories. In the first category are predictors that did not interact significantly with corpus type: subject NP head, post-head subject NP length, form of the copula, and stress clash. Of the predictors in this first category, stress clash deserves more lengthy discussion, because there is a numerical interaction between corpus type and stress clash that is marginal in statistical significance (p=0.06), with stress clash mattering less in writing than in speech,. More importantly, however, the effect of stress clash is highly significant in each corpus type individually (p<<0.001), with the same qualitative effect: potential stress clash favors *to*-use. Figure 9 illustrates this effect, with effect estimates and standard errors derived from a sum-coding representation of the interaction. The existence of the effect in the written data provides support for the part of Janet Fodor's (1998) "Implicit Prosody Hypothesis" that Fodor (2002) formulates as follows: "In silent reading, a default prosodic contour is projected onto the stimulus…" Moreover, it suggests that writers are influenced in their wording choices by this implicit prosody.

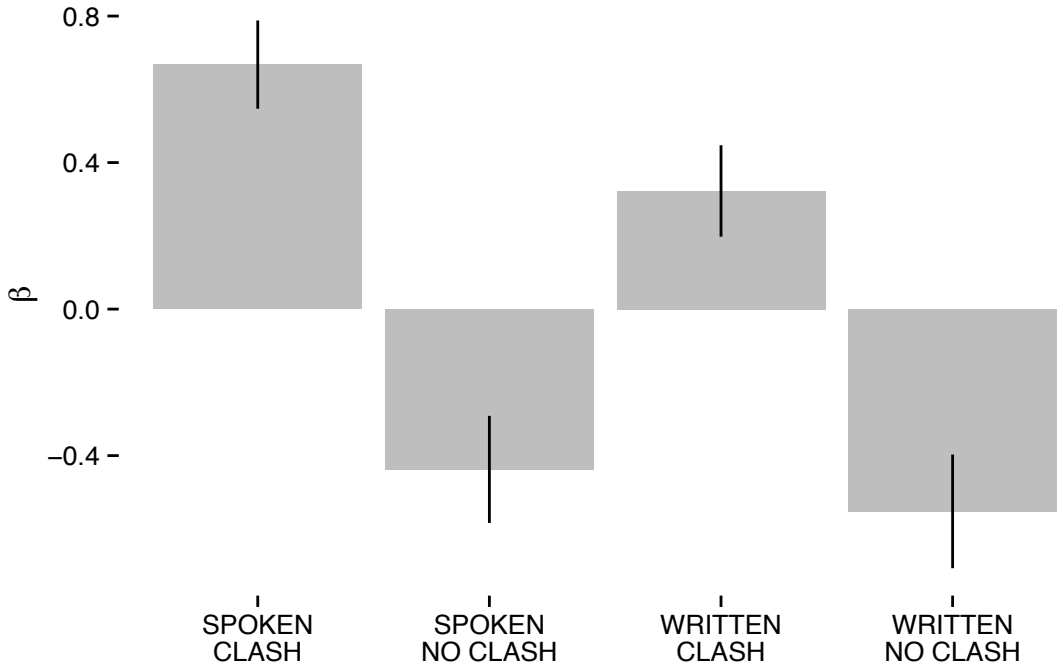

Figure 9: Effect of potential stress clash on *to*-use preference in speech and writing. Error bars show standard errors of the regression parameter estimates.

In the second category are predictors that interact significantly with corpus type, but not in ways that lead to a qualitative change in our overall picture: *do*-type, number of interveners between *do* and *be*, and PCVP length (all p<0.01). Number of *do*-*be* interveners and PCVP length have the same effects in speech and writing (with more of each favoring *to*-use), but in each case the effect is stronger in writing than in speech. In the case of *do*-type, the interaction involved the forms *did* and *done* favoring *to*-use more strongly in speech than in writing. These past-tense forms are less common in speech than in writing, so this result is also consistent with our theoretical picture of less frequent

components of the construction favoring *to*-use.

The sole predictor in the third category was the number of interveners between *be* and the PCV, which interacted significantly ($p<<0.001$) with corpus type in a theoretically important way. Recall that in the base model of Table 1, *be*-PCV interveners had no effect on *to*-use preference. However, adding an interaction with corpus type resulted in a far better fitting model (likelihood-ratio test $p<<0.001$). To understand this interaction we nested *be*-PCV interveners inside corpus type, and added random by-PCV slopes of *be*-PCV interveners and its interaction with corpus type; in this model, we found that more *be*-PCV interveners favored *to*-use in written English ($\beta=0.66$, $p<0.001$) but marginally disfavored *to*-use in spoken English ($\beta=-0.21$, $p=0.09$). A likelihood ratio test confirmed that the interaction between *be*-PCV interveners and corpus type is highly significant ($p<0.001$) in this model with maximal random effects structure with respect to this critical interaction (see Barr et al., 2013).

Why would the effect of *be*-PCV interveners, unlike all our other measures of utterance complexity, differ qualitatively between speech and writing? Consider this: additional material inserted between *be* and the PCV, unlike additional material in the NP subject or the postverbal part of the PCVP, is in the same position as optional *to*. While some types of *be*-PCV interveners may be semantically "full" – obligatory in order for the utterance to convey the speaker's intended meaning – others may be semantically "empty", and the speaker's use of them may be driven by the same considerations – utterance planning and prosodic optimization – that drive *to*-use. The following pair (both from the spoken part of COCA, italics indicate the intervener) illustrate this potential contrast, the first semantically "full" and the second "empty":

(10) (a) all we have to do is *not* continue the $100-billion-a-year increase that Obama and the Democrats put into domestic discretionary spending
(b) all it has to do is *just* jump down that hill right there

On this view, semantically "empty" material may sometimes be used instead of *to* and thus disfavor it. If this view is correct, and such semantically "empty" material disfavoring *to* is more common in speech than in writing, it could explain the discrepancy seen in the effect of *be*-PCV across the corpus types: a true underlying effect of semantically "full" material that favors *to* could be obscured by a higher incidence of semantically "empty" material in speech. We explored this hypothesis by focusing on the single most common *be*-PCV intervener, the word *just*. Although it is difficult to judge

	Spoken COCA	Written COCA
Singleton intervener is NOT *just*	34.6%	44.6%
Singleton intervener is *just*	14.7%	43.5%

Table 2: rate of *to*-use in speech versus writing for cases with a 1-word *be*-PCV intervener

when and to what extent *just* is semantically "full" versus "empty", there are few if any interveners that are likely to be "empty" more often than *just*. As it turns out, the

behavior of *just* is highly revealing. Table 2 shows the rate of *to*-use in speech and writing among utterances with single-word *be*-PCV interveners.[16] In writing, the rate of *to*-use is approximately the same for *just* and other single-word interveners. In speech, however, *just* disfavors *to*-use far more strongly than other single-word interveners.

This speech-specific dispreference of *just* for *to*-use provides initial confirmation of our hypothesis. We tested the hypothesis more rigorously by fitting a model with both the intervener by corpus type interaction, a main effect of single-word *just*, and an interaction between *just* and corpus type (with a maximal random effects structure with respect to these parameters). In this model, *just* significantly disfavored *to*-use in speech (β=-0.68, $p<<0.001$) but had no effect in writing (β=-0.01, p=0.96); more *be*-PCV interveners still favored *to*-use in writing (β=0.64, $p<0.005$) but now had no effect in speech (β=-0.03,p=0.78). That is, simply by accounting for the possible effect of *just* as behaving differently from other *be*-PCV interveners, the reverse effect of interveners in speech disappeared altogether. We speculate that the underlying effect of semantically "full" interveners may be to favor *to* in speech as in writing, but remains obscured by a longer tail of other semantically "empty" interveners individually less frequent that *just*. We leave assessment of this speculation as an open question for future research.

4. Conclusions and Directions for Future Work

Our study of optional *to* in the DBC suggests that processing factors familiar from the study of optional *that* play a major role in determining where *to* is used. These factors include measures of structural complexity and in-construction word frequency, including the specific prediction from the theory of Uniform Information Density that in-construction frequency of the post-copular verb will be negatively associated with *to* use. These findings support the idea that these factors apply quite generally to language production and are likely to influence the use of other optional function words in similar ways. We also found that prosody, a factor not included in models of optional *that*, seems to play an important role in determining whether speakers and writers use *to* before the post-copula verb in DBC sentences.

One broad theoretical consequence of our results is that they constitute evidence against a serial, modularist view of the lexical-selection and phonological-encoding stages of language production. Production is commonly seen as a cascaded process in which lexical selection precedes phonological encoding (Levelt, 1993). On a serial, modularist version of this view, preferences stated in terms of representations from the later stage of phonological encoding cannot affect decisions in the earlier stage; on an interactivist view, such effects are possible through self-monitoring and feedback (see Goldrick, 2006, and Jaeger et al., 2012, for discussion). Our evidence for prosodic effects on lexical selection favors the interactivist view.

A second broad theoretical consequence regards the nature of these interactivist effects. Our key empirical findings all involve the speaker making *to*-production decisions that

[16] In speech, 40% of these one-word interveners are *just*; in writing, the figure is 31%.

optimize the communicative properties of the utterance. These properties include the time available to prepare or recover from syntactically complex parts of the utterance, the information-density profile of the utterance, and the prosodic contour of the utterance. Our results thus support a view of moment-by-moment language production as being crucially guided by considerations of communicative optimality (Levy & Jaeger, 2007; Jaeger, 2010). Our results do not, however, speak directly to the familiar question of audience design (Clark & Murphy, 1982): do the effects we see on *to*-production reflect speaker-centric production pressures, or effort on the part of the speaker to optimize the utterance for the addressee? This question is beyond the scope of the present paper.

As another test of the generality of the influence of prosody on optional *to* use, we did a very preliminary check of *to* use in another construction where it is optional, namely, after the verb *help*. As the examples in (11) show, *help* can take VP complements that are either base or infinitival, irrespective of whether an object NP intervenes.

(11) a. a lot of people helped to find you
b. she has helped find dozens of people
c. it did help Austin to find her voice
d. he could help Luke find the gateway

We searched COCA for uses of the verb *help* followed by a verb, with or without an intervening personal pronoun. We did this separately for the spoken and written portions of the corpus.

We made the following working assumptions: *help* is normally stressed; *to* and personal pronouns in this position are typically unstressed; and a large majority of the verb tokens in our searches probably have initial stress[17]. Given these assumptions and the fact that both stress clash and stress lapse are disfavored, we expected to see a far higher rate of *to* use when no pronoun intervenes between *help* and the following verb. Including *to* after a pronoun puts two unstressed syllables next to each other, resulting in stress lapse. On the other hand, including *to* when no pronoun is present often prevents stress clash. Table 3 gives the results of these searches.

	Spoken Hits	Written Hits
HELP V	6989 (78%)	38000 (77%)
HELP *to* V	1957 (22%)	11225 (23%)
HELP PPRO V	5637 (88%)	22012 (90%)
HELP PPRO *to* V	746 (12%)	2578 (10%)

Table 3

[17] These assumptions need verification, and are deliberately stated with hedges. Obviously, many verbs are not stress-initial. But more frequent words tend to be shorter, so a high percentage of the verb tokens will be monosyllabic and hence stress-initial; and many polysyllabic verbs are also stress initial. The reasoning leading to our predictions does not go through when the pronoun gets contrastive stress, or when the form of *help* used is *helping*. But we are confident that our assumptions hold for enough of the data to make this a meaningful preliminary test.

In both speech and writing, the rate of *to* use after a personal pronoun is about half of what it is when no pronoun is present. This is what we predicted. Of course, the role of prosody in *to* use after *help* needs to be studied much more carefully, minimally by including further factors (like verb frequency), checking the actual stress patterns of the verbs, and by distinguishing between *helping* and the other (monosyllabic) inflections of *help*. But the pattern in Table 2 strongly suggests that prosody plays a role in the use of optional *to* after *help*, just as it does in the DBC. Moreover, the effect appears to hold in both speech and writing, providing additional support for Fodor's Implicit Prosody Hypothesis.

Returning to the DBC, while our study has made progress towards explaining why *to* is used where it is, a great deal of the variability remains unaccounted for. Our model indicates that individual post-copula verbs have different likelihoods of being preceded by *to*, over and above what can be explained by their in-construction frequencies. Assuming that these differences are not arbitrary lexical idiosyncrasies, we would like to discover what properties of verbs are associated with being preceded by *to* at higher rates.

We conjecture that verb semantics may be relevant, and we have begun investigating one semantic property, namely stativity. This based in part on a claim of Lakoff (1966), who used the DBC (with *what* as the subject head) as a diagnostic for non-stativity; that is, he claimed that stative verbs could not appear as PCVs in the DBC, giving examples like (12), which he prefixed with asterisks.

(12) a. What I did was hear the music.
b. What Harry did was know the answer.

Our dataset includes many counterexamples to Lakoff's categorical claim, for example (13).

(13) a. what we want you to do is hear some stories of the real-life people
b. one thing you need to do before you go in is know your rights

But Lakoff's claim was not entirely off base. He listed 28 stative and 28 non-stative verbs at the end of his paper, and a check of our dataset show that the non-stative ones occur in our collection at about six times the rate of the stative ones: 1378 for the non-statives (out of about 5 million total occurrences of these verbs in COCA) vs. 163 for the statives (out of about 4 million total occurrences of these verbs).

This suggests that there is a semantic incongruence between the DBC and stative predicates, which might make the combination harder to produce and comprehend. If so, this could lead to higher rates of *to* use in DBC examples with stative PCVs. Testing this requires some independent means of assessing the stativity of verbs. And since stative verbs have low frequency in the DBC, we will have to determine whether any effect of stativity on *to* use is already covered in our model by in-construction frequency. We are beginning to investigate these issues, but do not yet have results to report.

Much remains to be done before we know all the factors that influence the use of *to* in the DBC. And a true understanding of the phenomenon will require explanations of why

these factors influence *to* use as they do.

References

Agresti, A. (2002). *Categorical Data Analysis*. Wiley, second edition.

Anttila, A., M. Adams, and M. Speriosu. (2010) "The role of prosody in the English Dative Alternation". *Language and Cognitive Processes* 25(7/8/9), 946-981.

Baayen, R. H., D. J.Davidson, and D. M. Bates. (2008). "Mixed-effects modeling with crossed random effects for subjects and items". *Journal of Memory and Language*, 59(4):390–412.

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily (2013). "Random effects structure for confirmatory hypothesis testing: Keep it maximal". *Journal of Memory and Language*, 68(3):255–278.

Bresnan, J., A. Cueni, T. Nikitina, and H. Baayen (2007). "Predicting the dative alternation". In Boume, G., Kraemer, I., and Zwarts, J., editors, *Cognitive Foundations of Interpretation*, pages 69–95. Amsterdam: Royal Netherlands Academy of Science.

Cedergren, H. J. and S. Sankoff (1974). "Variable rules: Performance as a statistical reflection of competence". *Language*, 50(2):333–355.

Chambers, J. M. and T. J. Hastie (1991). "Statistical models". In Chambers, J. M. and T. J. Hastie (eds), *Statistical Models in S*, chapter 2, pages 13–44. Chapman and Hall.

Cieri, C., D. Miller, and K. Walker (2004). "The Fisher Corpus: a Resource for the Next Generations of Speech-to-Text". In *LREC* (Vol. 4, pp. 69-71).

Clark, H. H. (1973). "The language-as-fixed-effect fallacy: A critique of language statistics in psychological research". *Journal of Verbal Learning and Verbal Behavior*, 12:335–359.

Clark, H. H., & Murphy, G. L. (1982). "Audience design in meaning and reference." In J. F. L. Ney & W. Kintsch (Eds.), *Language and comprehension* (Vol. 9, pp. 287–297). North Holland.

van Draat, P. F. (1910) *Rhythm in English Prose*. Carl Winter's Universitätsbuchhandlung. Heidelberg.

Flickinger, D. and T. Wasow (2013) "A Corpus-driven Analysis of the Do-Be Construction". In P. Hofmeister and E. Norcliffe (eds) *The Core and the Periphery: Data-Driven Perspectives on Syntax Inspired by Ivan A. Sag*, 35-63. CSLI Publications.

Fodor, J. D. (1998) "Learning to parse?" *Journal of Psycholinguistic Research* 27:285-319.

Fodor, J. D., 2002. “Prosodic disambiguation in silent reading”. *Proceedings of NELS 32*, M. Hirotani (ed.). Amherst, MA: GLSA, University of Massachusetts.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language & Cognitive Processes*, 21(7–8), 817–855.

Hawkins, John A. (1994) *A Peformance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. London: Chapman & Hall.

Jaeger, T. F. 2010. “Redundancy and Reduction: Speakers Manage Syntactic Information Density”. *Cognitive Psychology*, 61(1), 23-62.

Jaeger, T. F., Furth, K., & Hilliard, C. (2012). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 38 (5), 1439–1449.

Klein, D. and C. D. Manning (2003) “Accurate Unlexicalized Parsing”. *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430.

Lakoff, G. (1966) “Stative Adjectives and Verbs in English”. In A.G. Oettinger (ed) *Mathematical Linguistics and Automatic Translation*. Report NSF 19, Computation Laboratory, Harvard University.

Levelt, W. J. M. (1993). *Speaking: From intention to articulation*. MIT Press.

Levy, R. P., and T. F. Jaeger (2007) “Speakers optimize information density through syntactic reduction”. In *Advances in neural information processing systems* (pp. 849-856).

Liberman, M. and A. Prince (1977) “On stress and linguistic rhythm”. *Linguistic Inquiry* 8. pp. 249–336.

Pinheiro, J. C. and D. M. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. Springer.

Rohde, Douglas L. (2005) “Tgrep2 User Manual”. http://tedlab.mit.edu/~dr/Tgrep2/tgrep2.pdf

Shannon, C. E. (1948) “A mathematical theory of communications”. *Bell Systems Technical Journal*, *27*, 623-656.

Warren, T. and Gibson, E. (2002). The influence of referential processing on sentence complexity. Cognition, 85(1):79–112.

Wasow, T. (2002) *Postverbal Behavior*. Stanford, CA: CSLI Publications.

Wasow, T., R. Greene, and R. Levy, "Optional *to* and Prosody". Poster at the 25th annual CUNY Conference on Human Sentence Processing. New York, March 2012.

Wasow, T., T. F. Jaeger, and D. Orr (2011) "Lexical Variation in Relativizer Frequency". In H. Simon and H. Wiese (eds) *Expecting the Unexpected: Exceptions in Grammar*. De Gruyter. 175-195.

Zipf, G. (1936) *The Psychobiology of Language* (Routledge, London).