Studying syntactic priming in corpora

Implications of different levels of granularity

Stefan Th. Gries
University of California, Santa Barbara

This chapter addresses syntactic priming (of the dative alternation) using corpus data from the ICE-GB corpus. Nearly 3,000 consecutive prime-target pairs were coded for their constructional choices as well as several other variables. The data are then analyzed on different levels of granularity: (i) cross-tabulation, (ii) with a binary logistic regression (including only fixed effects), and (iii) a generalized linear mixed-effects model (GLMEM) (including fixed and random effects). The GLMEM reduces the number of significant predictors most, but nevertheless yields the highest classification accuracy. Since contemporary cognitive linguistics assumes an item-based perspective on language acquisition, processing, and change, I argue that the GLMEM approach should be the method of choice for empirical cognitive linguistics.

Keywords: binary logistic regression models, corpus data, cross-tabulation, dative alternation, generalized linear mixed-effects models, item-specificity, empirical cognitive linguistics

1. Introduction

Syntactic priming/persistence, the tendency of speakers to re-use syntactic patterns they have recently comprehended or produced, is a phenomenon that has attracted considerable attention ever since Bock's pioneering work in the early 1980s. In a series of publications, Bock and colleagues showed that speakers who read (1a) are more likely to describe a transitive scenario with a passive sentence than speakers who read (1b).

(1) a. *The duckling was killed by the farmer.*
  b. *The farmer killed the duckling.*

Much initial work on this tendency focused on demonstrating that such priming effects are in fact a tendency to use identical syntactic constituent structures rather than, say, identical metrical structures or identical thematic role orderings. For example, Bock and Loebell (1990) showed that (2a) primes prepositional datives such as *John [sent [$_{NP}$ a book] [$_{PP}$ to [$_{NP}$ Mary]]]*, but (2b) does not even though it has the exact same metrical structure (but not the relevant syntactic structure):

(2) a. *Susan brought a book to Stella.*
 b. *Susan brought a book to study.*

Bock (1989) showed both (3a) and (3b) prime passive such as *[$_{PAT}$ The duckling] was killed [$_{PP}$ by [$_{AGT}$ the farmer]]* even though only (3a) involves a *by*-passive:

(3) a. *The 747 was alerted by the airport's control tower.*
 b. *The 747 was landing by the airport's control tower.*

Syntactic priming has turned out to be a very general and robust effect. It has been obtained

– from production to production (cf. Bock 1986);
– from comprehension to production (cf. Branigan et al. 2000, Bock et al. 2007);
– when the verb lemmas in primes and targets are the same or different (cf. Pickering and Branigan 1998);
– in L1 (cf. all studies quoted so far) and L2 (cf. Bock and Loebell 2003, Gries and Wulff 2003, 2009);
– across different languages (cf. Hartsuiker, Pickering, and Veltkamp 2004, Schoonbaert, Hartsuiker, and Pickering 2007, and Shin and Christianson 2009).

With very few exceptions, the work on syntactic priming has been experimental in nature, involving many different kinds of paradigms – studies involving corpus-based approaches are few and far between (cf. Estival 1985, Gries 2005, or Szmrecsanyi 2005). As a matter of fact, it is not uncommon to hear or read statements that question the utility of corpus data in the study of syntactic priming; the following is a case in point.

> [T]here are several nonsyntactic factors which could lead to repetition. [...] Corpora have proved useful as a means of hypothesis generation, but unequivocal demonstrations of syntactic priming effects can only come from controlled experiments (Branigan et al., 1995 492; cf. also Pickering and Branigan, 1999: 136).

In this study, I argue that, in spite of a degree of noisiness that exceeds that of experimental data, corpus data can not only coincide strikingly with experimental data – and, thus, provide converging evidence – but also have things to offer to priming researchers that are interesting and hard to obtain with the usual carefully-controlled experimental designs. More specifically, based on a reanalysis of data discussed in Gries (2005), I show that corpus data have a lot to add to the customary experimental approaches, in particular when it comes to (i) the analysis of the duration of priming and (ii) the issue of how priming effects are specific to the verbs involved and to (at least approximately) individual speakers/authors.

As for the duration of priming effects, there are conflicting findings. On the one hand, some studies show that the distance between prime and target has to be fairly small in order for priming effects to be obtained (cf. Levelt and Kelter 1982 or Branigan et al. 1999). On the other hand, other studies show that priming may well persist for longer times and across quite a bit of intervening material (cf. Bock and Griffin 2000, Pickering et al., 2000, Chang et al., 2000, and Bock et al. 2007). However, given the large number of distances and types of intervening material between primes and target, it is difficult to assess very many different distances between primes and targets in carefully controlled experimental settings. In addition, it is well-known by now that not all syntactic constructions are equally sensitive to priming effects, which brings us to the next topic, lexical specificity.

Previous studies have shown that different constructions are differently responsive to priming both 'between alternations' and 'within alternations'. As for the former, the dative alternation exhibits stronger priming effects than the voice alternation (cf. Bock 1986: Exp. 1); as for the latter, ditransitives are primed more strongly than prepositional datives (cf. Bock 1986; cf. Potter and Lombardi 1998 for the opposite result). But what about the lexical material – more precisely, the verb – that is involved in the clauses tested for priming? Not only is it well known that verbs have differently strong probabilistic preferences to occur with particular constructions, it is also well known that these preferences correlate with many other aspects of psycholinguistic processing such as ease and speed of lexical access or ambiguity resolution (cf. Garnsey et al. 1997, Stallings et al. 1998, Hare et al. 2003). It is therefore reasonable to assume that, if verbs 'prefer' to occur in constructions differently strongly, then they will be differently susceptible to priming effects. Surprisingly, this possibility has rarely been mentioned in experimental studies; see Potter and Lombardi (1998: 278) for an exception, and most studies have been content to provide F_1/F_2 and quasi-F statistics on the data to determine whether observed effects are significant across the ranges of verbs and speakers included in the experiment. However, as I already pointed out, not all studies chose their experimental stimuli such that they systematically and symmetrically exhaust the whole range of (strengths of) subcategorization preferences. And to some degree that is understandable:

- the number of verbs that participate in the usual suspects of alternations can be so large as to make it impossible to include a larger share of them into an experiment, especially when all other independent variables and the common experimental controls are not just included but systematically and exhaustively varied and crossed.
- operationalizing/measuring subcategorization preferences is not a straightforward matter. Most studies so far have used raw observed frequencies (cf., Connine et al. 1984 or Hare et al. 2003), but corpus-based research shows that it may be much more useful to use measures that control for verbs' and constructions' overall frequencies in a corpus (cf. Gries, under review).

More recently, in the study of alternation patterns, Gries and Stefanowitsch (2004) proposed the measure of distinctive collexeme strength to quantify a verb's preference for one out of several functionally similar constructions. This measure is typically based on computationally somewhat intensive Fisher-Yates exact tests of 2×2 co-occurrence tables and provides a kind of estimate of verbs' preferences to one of two constructions. Some experimental studies have now shown that this measure is a strong predictor of subjects' verb-specific construction preferences in priming tasks (Gries and Wulff 2003, 2009) and can even outperform the usual kind of raw frequencies in sentence completion and self-paced reading tasks (cf. Gries, Hampe, and Schönefeld 2005, 2010). A rigorous corpus-based approach thus not only provides a better operationalization of individual verbs' constructional preferences, but also the opportunity to include very many different verbs and their preferences into the analysis of structural priming effects; ultimately, this affords to the analyst a versatile and powerful tool with which to obtain (con- or diverging) evidence difficult to attain otherwise.

In the remainder of this chapter, I reanalyze corpus data on the dative alternation first studied in Gries (2005), but I go beyond that study by comparing three different ways to analyze the corpus data statistically. The first of these ways, the study of mere constructional frequencies, provides just a very simple lowest level of sophistication. The second of these ways – a binary logistic regression analysis – provides a much more precise set of results. Most importantly, however, the third way is the statistical approach of generalized linear mixed-effects modeling, a new technique which not only has several statistical advantages over more traditional methods but is also more compatible with the degree of importance attached to item-based patterns by usage-based linguists these days; these issues and their implications are discussed in more detail below.

The following section introduces in more detail the data that were investigated in this study and the analytical tools used.

2. Data and methods

In order to investigate syntactic priming corpus-linguistically, I first identified ditransitive constructions and prepositional datives with *to* and *for* in the British component of the International Corpus of English (ICE-GB).[1] These data were cleaned such that clauses that were the first or last construction either in one of the 500 corpus files or in a subtext of a corpus file were discarded (because they cannot function as targets or primes respectively) leaving 2,877 prime-target pairs (i.e. subsequent construction pairs of either type) for the analysis that could be coded for all intended variables.

The variables that are studied included both fixed and random effects. The dependent variable was CTarget, the construction of the second of the two constructions constituting a prime-target pair: *ditransitive* vs. *prepositional dative* (automatically retrieved from the annotated parse trees within the corpus files). The fixed effects I included were the following:

- CPrime: the construction of the first of the two constructions constituting a prime-target pair: *ditransitive* vs. *prepositional dative* (automatically retrieved from the annotated parse trees within the corpus files);
- Medium: the medium in which prime and target occurred: *spoken* vs. *written* (automatically retrieved from the corpus files);
- LogDistance: the distance in parsing units between the occurrence of prime and target within each subtext of each file as determined from the annotation of the corpus, which was logged and centered;
- VFormID and VLemmaID: whether both constructions involved the same verb form and verb lemma: *yes* or *no* (cf. Pickering and Branigan 1998);
- SpeakerID: whether in the spoken data both constructions were produced by the same speaker or not: *yes* or *no*.

In addition to these main effects, I also included the interactions of CPrime with all other independent variables. This means, for example, that I checked whether the effect of the construction in the prime is differently strong in speaking and writing

1. The ICE-GB is a POS-tagged and fully parsed corpus of spoken and written British English of the 1990s; all annotation has been checked manually by several linguists (cf. http://www.ucl.ac.uk/english-usage/ice-gb/index.htm for details). The data discussed here differ slightly from those used in Gries (2005) because VPs with *to*/*for* were extracted using different fuzzy tree fragments and the distinctive collexeme strengths were updated accordingly.

(the first interaction), whether the effect of the construction in the prime varies as a function of the distance between prime and target (the second interaction), ...[2]

The random effects, on the other hand, were as follows:

- VLemmaTarget: the exact verb lemma of each target;
- File: the name of the file in which the prime and target were observed.[3]

After all prime-target pairs were coded for all these variables, I performed the three statistical analyses briefly mentioned above. These analyses differ in terms of their level of granularity and statistical sophistication in the sense that they use successively finer degrees of resolution in their characterization of the situation that gives rise to a particular constructional choice.

First, at the coarsest possible level of granularity or precision, the only independent variable included is CPrime. That is, one just cross-tabulates the constructional frequencies in prime and target slots, which obviously provides only a very crude measure of whether priming can be observed and, if so, how strong the priming effect is.

Second, at a finer level of granularity, more information about the situation at the time of production is included. More specifically, one can include all

2. It is worth pointing out in this context how binary logistic regression goes beyond what a traditional variationist tool such as Varbrul can do: not only can such regressions handle continuous data, they can also include interactions of factors seamlessly; cf. <http://www.ling.upenn.edu/~johnson4/Rbrul_manual.html> for an R function that offers these kinds of functionality as well as the kind of random effects to be discussed below.

3. One reviewer raised the question of why verbs are not entered into the GLMEM analysis as a fixed effect, arguing that verb lemma effects can be considered "repeatable" (quoting Baayen 2008: 141) and that, therefore, modeling verb-specific effects as a random effect is more a matter of statistical convenience than theoretical conviction. I agree that modeling verb-specific effects as random effects is methodologically more convenient and that, in the spirit of Gries and Stefanowitsch's (2004) collostructional analysis, verb-specific effects can be seen as repeatable.

On the other hand, note that even the reference quoted by the reviewer – Baayen (2008), who refers to repeatability in his discussion of fixed and random effects – uses random effects to model item-specific effects in corpora, and this seems to be the (currently emerging) standard of how mixed-effects models are used in linguistics (cf. also Johnson's (2008: sect. 7.3–7.4) discussion). In addition, another criterion often referred to in this connection has to do with whether the levels of a factor in the sample that is being studied exhaust the full range of levels this factor has in the population. In the case of factors like sex of subject (in experiments) or CPrime (here), this is the case, and such factors are typically modeled as fixed effects. However, the verb lemmas studied in this paper's sample do not exhaust all possible verbs that can be used ditransitively in the population, which supports the inclusion of VLemmaTarget as a random effect. Finally, recall that GLMEM are better at taking different frequencies of factor levels and group-level variability into account than regular regression models (cf. Luke (2004: 6–7) as well as Gelman and Hill (2007: 245–246) for more discussion).

above-mentioned fixed effects and (some of) their interactions and perform a generalized linear model analysis on the data, in which the choice of construction in the target slot is predicted on the basis of these predictors. In Gries (2005), I performed an analysis of that kind, but used an ANOVA, which, with hindsight, was a sub-optimal methodological choice (cf. Jaeger 2008 for discussion of shortcomings of ANOVAs). In the second analysis of this study, I follow the exemplary study of Szmrecsanyi (2005) and use a binary logistic regression, which is the more appropriate measure since it does justice to the facts that the dependent variable is categorical and its distributional assumptions are more compatible with the data.

Finally, and most importantly for the present study, one can describe the situation at the time of production with an even higher degree of granularity. In this chapter, I am referring to the possibility of performing a generalized linear mixed-effects model (GLMEM), i.e. a binary logistic regression including not only the above fixed effects but also the random effects. That is, a part of the variation within the data is accounted for by including random effects for files (as a rough (!) approximation to distinguishing different speakers/writers, which in experimental studies is often captured by by-subjects analyses) and for verb lemmas in the target (as an operationalization of verb-specific preferences, which are often captured by by-items analyses). This approach has several advantages over more traditional methods.

First, GLMEM is geared towards including random effects such as, here, verb-specific effects and corpus file-specific overall patterns. While both of these kinds of random effects were in fact discussed in Gries (2005), verb-specific effects were only included by discussing a handful of verb preferences (experimental items used in Pickering and Branigan 1998), and corpus file-specific effects were only explored graphically using switch-rate plots. In this study, using GLMEM, I can include these two kinds of effects in a statistically more comprehensive way.[4] This has the advantage that, while these random effects are not predictors of the constructional choice *per se*, they make the estimation of the predictors more precise and more robust, and I compare binary logistic regression and GLMEM to show why this is important and revealing. Second, such mixed-effects models are better than traditional regression methods at handling the fact that, for example, different verbs will be differently frequent in the data and will, therefore, contribute different amounts of information to the model (cf. Gelman and Hill 2007: 246, 254). Finally, Baayen (2008: Section 7.2.1) shows how mixed-effects models outperform the current standard in psycholinguistics of reporting F_1/F_2 and quasi F statistics or ANOVAs on transformed data (cf. again Jaeger 2008).

4. The included random effects are only adjustment to intercepts, not also to slopes.

Before we turn to the discussion of the results in the following section, one final caveat is in order. Quite obviously, there are quite a few additional variables that could be included in the analyses to be discussed below. For example, Bresnan et al. (2007) include information-structural variables and constituent length differences in their study (and even verb sense as a random effect). Also, Jaeger and Snider (2007) include a different kind of verb-specificity effect called surprisal and show that, if a prime consists of a verb in a construction that the verb is not generally associated with, then the priming effect is stronger. It goes without saying that the inclusion of more independent variables and/or random effects will not only increase the overall classification accuracy but also increase the probability that what is studied are really priming effects of CPrime and not other factors facilitating structural repetition. However, the classification of approximately 6,000 recipient and patients for, say, discourse accessibility is beyond the scope of the current chapter since the focus here is not on demonstrating that corpora can be used to study the psycholinguistic process of priming – some of the studies mentioned above have done that on a larger and more comprehensive scale than is possible here – but on showing that

- different quantitative approaches to corpus data differ substantially in terms of (i) the determinants of structural repetitions they identify, (ii) as a corollary, the quality of the results they yield, and (iii) their theoretical fit to the usage-based commitment of much of contemporary cognitive linguistics;
- the most appropriate statistical analysis of corpus data yields results supportive of, and complementary to, experimental data, which ultimately, supports the idea of converging evidence from methodological pluralism.

3. Results

3.1 Coarse granularity: Constructional frequencies

The first type of analysis at the coarsest level of granularity results in the observed frequencies in Table 1.

Table 1. Observed frequencies of both constructions in primes and targets

	CTarget: prep. dative	**CTarget: ditransitive**	**Totals**
CPrime: prep. dative	746	549	1,295
CPrime: ditransitive	514	1,068	1,582
Totals	1,260	1,617	2,877

Table 2. Expected frequencies of both constructions in primes and targets

	CTarget: prep. dative	CTarget: ditransitive	Totals
CPrime: prep. dative	607.7	687.3	1,295
CPrime: ditransitive	742.4	839.6	1,582

These observed frequencies must be compared to the frequencies expected by chance. These are computed here on the assumptions that (i) after each construction in CPrime, the speaker can produce either construction in the target, and (ii) the random, baseline probabilities to choose one construction in CTarget follows from the constructions' overall frequencies in the corpus:

Once this pattern is represented graphically (as in Figure 1), the result becomes clearer.

There is a marked priming effect: when the prime was a prepositional dative, then prepositional datives and ditransitives were more and less frequent than expected respectively, and when the prime was a ditransitive, then prepositional datives and ditransitives were less and more frequent than expected respectively. Note also that the strengths of the preferences for the two constructions are very similar to those reported in Bock's pioneering experimental study. Bock (1986: 364) found percentage ratios of prepositional dative and ditransitive preferences of 1.5 and 2.1 respectively (i.e. an 'odds ratio' of approximately 0.71), while I obtained 1.45 and 1.95 respectively (i.e. an 'odds ratio' of approximately 0.75). In spite of many potential additional predictors ignored in this study, this points to a large degree of convergence of the experimental and the corpus-based approaches.

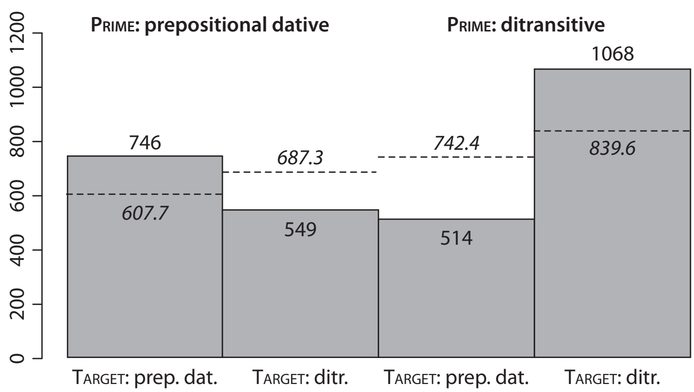


Figure 1. Bar plot with observed construction frequencies (with horizontal lines and italic numbers indicating expected frequencies)

While this result is straightforward to interpret, it is also not particularly informative since it does not provide any more detailed information. Here, clearly, a more refined method of analysis is required.

3.2 Intermediate granularity: Binary logistic regression

The statistical analysis on the intermediate level of granularity supports the much more shallow analysis of Section 3.1, but provides a wealth of additional information. Starting out from the maximal model that includes all above fixed effect predictors (main effects and interactions), I went through a model selection process that in a stepwise fashion (from interactions to main effects) eliminated all predictors that were not significant and did not participate in a higher-order significant interaction. The final, minimally adequate model shows that there is a highly significant correlation between the fixed-effect predictors that survived the model selection process and the construction chosen in the target (model L.R. $\chi^2 = 1340.76$; $df = 8$; $p < 0.001$). This final model has a reasonable degree of classificatory/predictive power ($C = 0.718$; Somer's $D_{xy} = 0.436$; Nagelkerke's $R^2 = 0.248$); cf. Table 3 for its classification matrix.

According to the chance expectation – always choosing the more frequent construction – one would get $^{1,617}/_{2,877} = 56.2\%$ correct classifications. The classification accuracy obtained in the model, however, is somewhat better: $^{1,833}/_{2,877} = 63.71\%$, and the difference between these percentages is highly significant ($p_{binomial\ test} < 0.0001$). It is worth briefly comparing the present classification accuracy to that of, say, Bresnan et al. (2007). On the one hand, the present classification accuracy is much worse because they achieved classification and prediction accuracies of 92%. On the other hand, this was to be expected, given that they included a much larger number of predictors in their model. Potentially more importantly, the performance can maybe be evaluated more accurately by considering the baseline accuracies of both studies. Bresnan et al.'s model improved the accuracy from the baseline of 79% to 92%, i.e. by approximately 16%, while the present model improved the accuracy from the baseline of 56.2% to 63.7%, i.e. by approximately 13%, which is much closer to Bresnan et al.'s improvement than a mere comparison of classification accuracies would suggest.

Table 3. Classification matrix of the binary logistic regression classifications (with italic numbers indicating correctly predicted constructions)

predicted/observed	**CTarget: prep. dative**	**CTarget: ditransitive**	**Totals**
CTarget: prep. dative	*720*	504	1,224
CTarget: ditransitive	540	*1,113*	1,653
Totals	1,260	1,617	2,877

The question remains, what are the predictors that made it into the final model and how strongly do they affect the constructional choice. Figure 2 represents the relevant predictors and how they influence CTarget. On the *x*-axis, I show all predictors plus the level whose coefficient is represented in the graph. Against the *y*-axis, I represent the coefficient size (with the "×") and its standard error (with the error bars). Positive and negative coefficients indicate preferences for the prepositional dative and the ditransitive respectively, and the more a coefficient deviates from zero, the stronger the effect; significance levels are indicated at the top of the graph.

As can be seen on the left of Figure 2, the main effect of CPrime is not significant (cf. the "ns" at the top and the fact that the standard error includes 0) and other main effects are significant but irrelevant for the study of priming (the fact that the proportion of ditransitives is larger in writing does not really mean much in this context). However, most main effects participate in interactions with CPrime; cf. Figure 3.

The first interaction shows that prepositional dative primes tend to yield prepositional dative targets, but that this effect is very much and highly significantly stronger when the verb form in the target is also exactly the same as in the prime. The second interaction shows that the same pattern holds with regard to whether the same verb lemma is used in both prime and target. The third interaction shows that prepositional dative primes tend to yield prepositional dative targets significantly more strongly when the speaker/writer is the same across both and target.

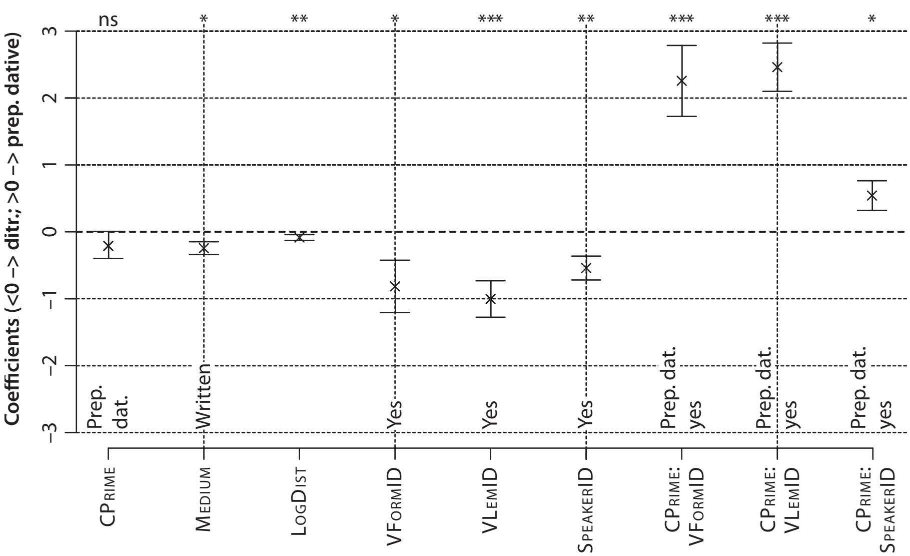
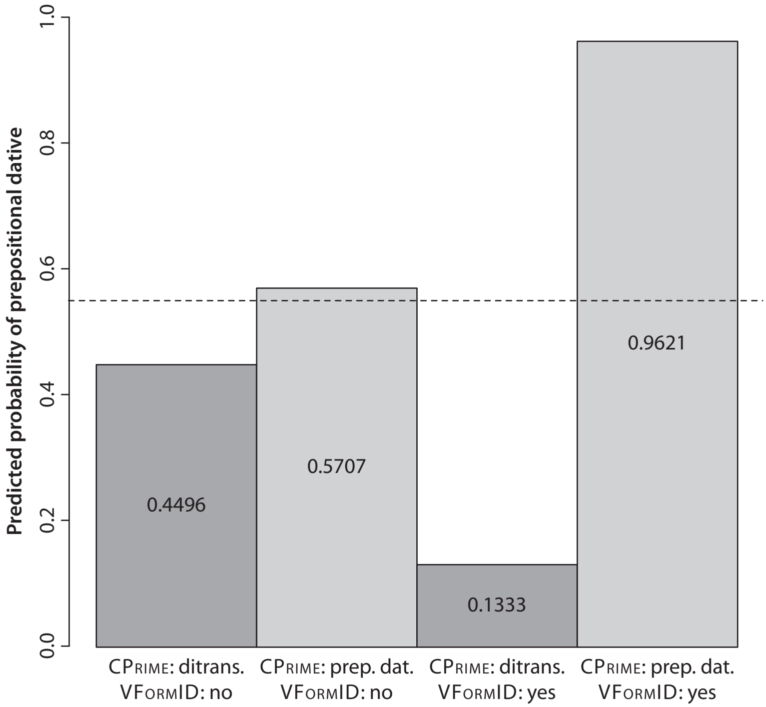
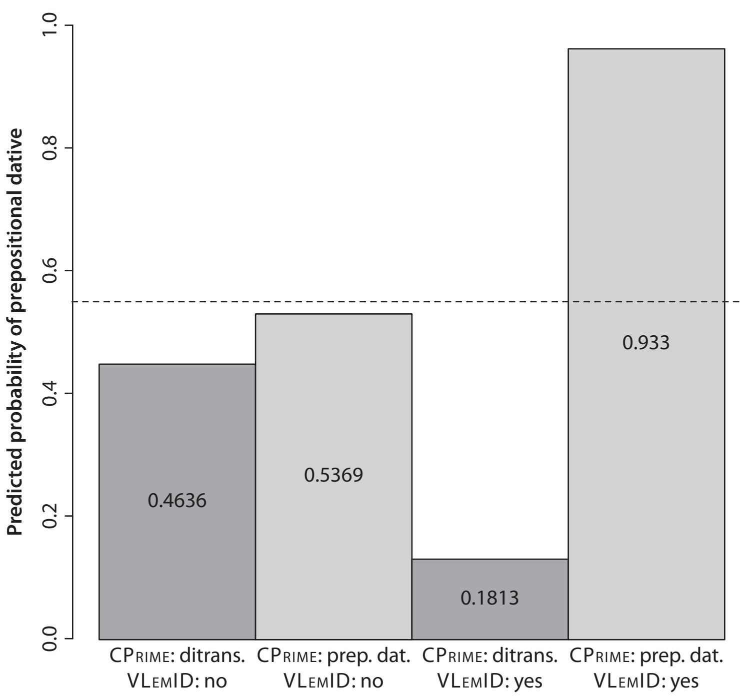


Figure 2. Coefficients of the independent variables in the minimal adequate model of the binary logistic regression (with their standard errors)


Predicted probability of prepositional dative
0.0
0.2
0.4
0.6
0.8
1.0
0.4496
0.5707
0.1333
0.9621
CPrime: ditrans. VFormID: no
CPrime: prep. dat. VFormID: no
CPrime: ditrans. VFormID: yes
CPrime: prep. dat. VFormID: yes


Predicted probability of prepositional dative
0.0
0.2
0.4
0.6
0.8
1.0
0.4636
0.5369
0.1813
0.933
CPrime: ditrans. VLemID: no
CPrime: prep. dat. VLemID: no
CPrime: ditrans. VLemID: yes
CPrime: prep. dat. VLemID: yes

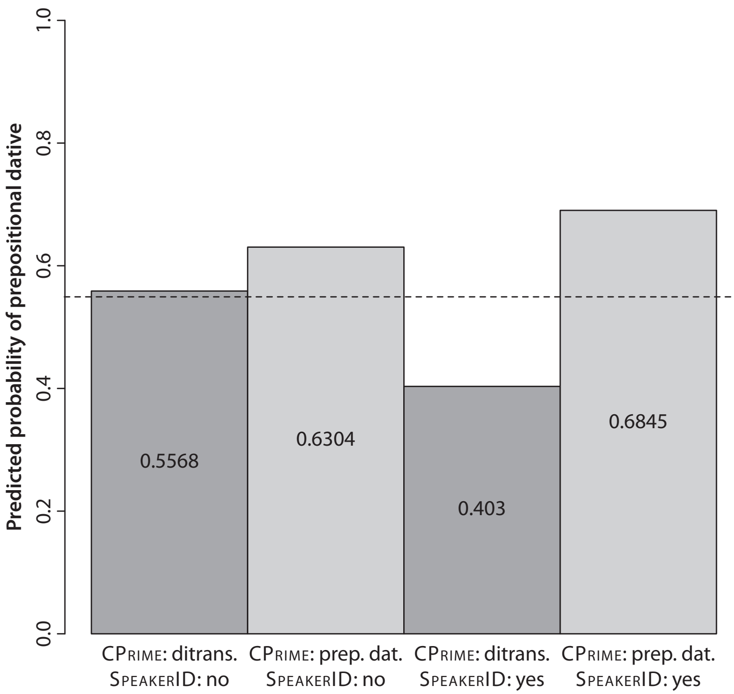


Figure 3. Interactions of VFormID, VLemmaID, and SpeakerID with CPrime (with the predicted probability of prepositional datives on the y-axis)

In sum, the data in general confirm some previous work cited above: syntactic priming exists and it is stronger the more prime and target are similar to each other otherwise, be it the verb form, the verb lemma, and/or the speaker. In the following section, the corpus data are explored even more precisely.

3.3 High granularity: Generalized linear mixed-effects model

The final analysis in this section is somewhat more complex and I can discuss neither all technicalities here (cf. Gelman and Hill 2007, Baayen 2008: Ch. 7, Johnson 2008: Sect 7.3–7.4) nor devote much time on the conceptual implications – instead, the focus is on exemplifying how the results that a GLMEM approach provides go beyond the standard binary logistic regression; the GLMEM was computed with the function lmer in the R environment (cf. R Development Core Team 2009).

Again, I started out from the maximal model that includes all fixed effect predictors (main effects and interactions) from above plus, now, two random effects as adjustments to the intercept. Then I first went through the same kind of model selection process, eliminating in a stepwise fashion (from interactions to main

Table 4. Classification matrix of the GLMEM regression classifications (with italic numbers indicating correctly predicted constructions)

predicted/observed	CTarget: prep. dative	CTarget: ditransitive	Totals
CTarget: prep. dative	*1,084*	117	1,201
CTarget: ditransitive	176	*1,500*	1,676
Totals	1,260	1,617	2,877

effects) all predictors that were not significant and did not participate in a higher-order significant interaction. In a second step, I tested whether each random effect could be removed with a significant loss of information, which was not the case.

The final, minimal adequate model shows that there is again a strong correlation between the fixed-effect predictors that survived the model selection process and the construction chosen in the target. This final model has a very high degree of classificatory/predictive power, as is indicated in its classification matrix in Table 4.

According to the chance expectation – always choosing the more frequent construction – one would again get $^{1,617}/_{2,877}$ = 56.2% correct classifications. The classification accuracy obtained in the random effects model, however, is extremely high, given that we are looking at noisy observational data in the behavioral sciences: $^{2,584}/_{2,877}$ = 89.82%. This classification accuracy is of course significantly better than the chance expectation – after all, the binary logistic regression was already better – but also highly significantly better than the traditional binary logistic regression (both $p_{binomial\ test} < 0.0001$) and remarkably close to the 92% and 94% accuracies that Bresnan et al. (2007) obtain on the basis of about three times as many predictors.

Again, what are the predictors that made it into the final model and how strongly do they affect the constructional choice? Figure 4 represents the relevant predictors and how they influence CTarget in a similar way as in Figure 2.

In some sense, the result is striking because the excellent classification accuracy is arrived at with only three fixed-effect predictors: CPrime, VFormID, and, crucially, their interaction plus the two random effects. In this model, prepositional datives in the primes already increase the chance of prepositional primes in the target (the main effect), but again the interaction CPrime:VFormID shows that this effect is particularly strong; the corresponding barplot looks very much like the left panel of Figure 3. None of the other predictors from the previous section reaches standard levels of significance (not even LogDistance, which is compatible with Bock et al. 2007: experiment 1). The question, thus, arises where the drastic 25% improvement of the classification accuracy comes from.

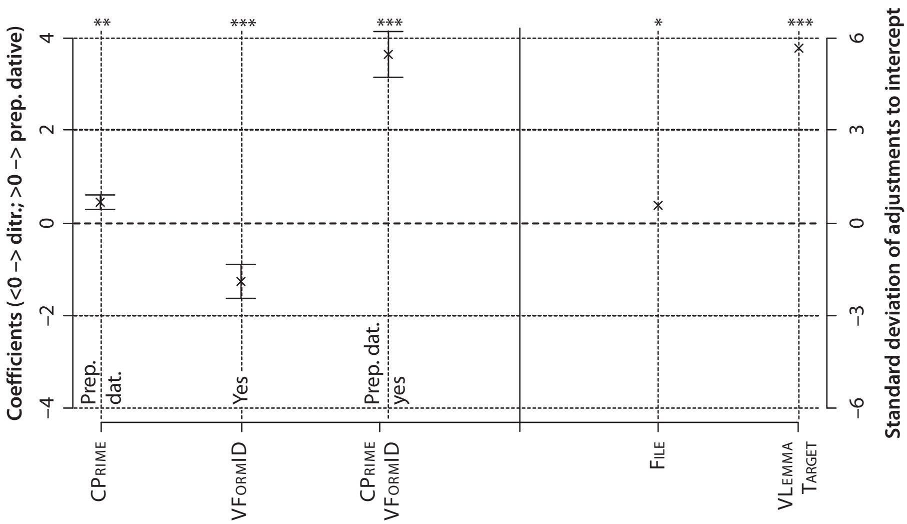


Figure 4. Fixed and random effects in the minimal adequate model of the GLMEM (left y-axis: fixed effects with standard errors; right y-axis: random effect standard deviations)

On the one hand, the improvement results from the fact that the coefficients of the regression are slightly different because the model fitting process took the different frequencies of verb lemmas and files into consideration and adjusted the coefficients accordingly. On the other hand and more importantly, the improvement results from the fact that every file and every verb lemma in the target now have their own regression intercept, which of course also affects – i.e. improves – the classifications. The most dramatic impact of these file/verb-specific intercepts is that they change the probability of the predicted construction – the prepositional dative – and when that predicted probability was below 0.5 in the binary logistic regression but increased to over 0.5 in the GLMEM, then the models make different classifications, and apparently the GLMEM is much more successful.

For example, there were 888 cases where the GLMEM made the right classification, but the binary logistic regression did not. Of these cases, 38 involved the verb lemma *send*, which was the lemma in a target construction 151 times. In other words, with the GLMEM, $^{38}/_{151}\approx 0.25$ = 25% of the verb lemma *send* are predicted better than with the logistic regression. Similar computations can be done for each verb, which can then be represented summarily as in Figure 5.

Overplotting of verbs notwithstanding, it becomes obvious immediately why a verb-specific account is so important: the verb lemmas differ drastically in terms of how much their classifications vary as a function of whether their idiosyncratic

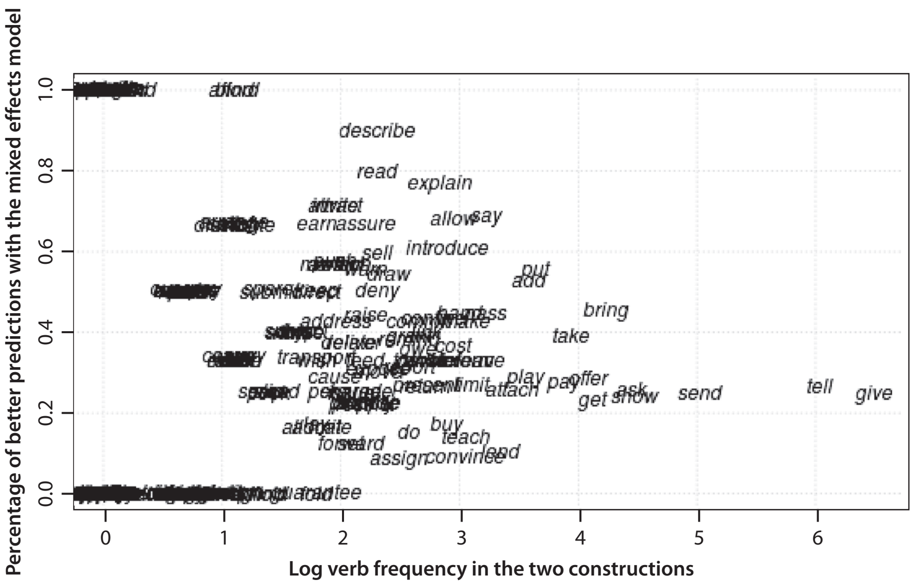


Figure 5. Percentage of improved classifications for each verb (on the y-axis) against each verb's logged frequency (on the x-axis)

preferences are included or not. Verbs such as *guarantee* do not benefit at all from a mixed-effects model approach; some of the higher-frequency verbs such as *give, tell*, *send*, *show*, *take*, and others benefit intermediately much, but some verbs cannot be handled by a regular logistic regression approach at all because (nearly) all their correct classifications only arise in the GLMEM model: *read*, *describe*, *afford*, *accord*, and others overplotted in the top left corner.

The same logic can of course be applied to the file-specific variation. I do not show the resulting plot here since the file names *per se* are not particularly revealing, but it is important to note in this connection how the inclusion of verb-/file-specific variation as random effects in a GLMEM goes beyond the kind of switch rate scatterplots proposed by Sankoff and Laberge (1978). Sankoff and Laberge plot the switch rates to a construction per file/speaker (on the *y*-axis) against the proportion of that same construction per file/speaker (on the *x*-axis), and when most dots are below the main diagonal, switches from one construction to the other (i.e. the absence of priming) are rarer across files/speakers than chance would predict (cf. Gries 2005: 395–6). While these plots are therefore a good visual diagnostic for problematic patterns in the data – e.g., files/speakers that exhibit priming effects or their absence that deviate from the norm considerably – the information they provide is not also used in the process of the statistical modeling. Put differently, if a particular file or speaker appears to be an outlier, then

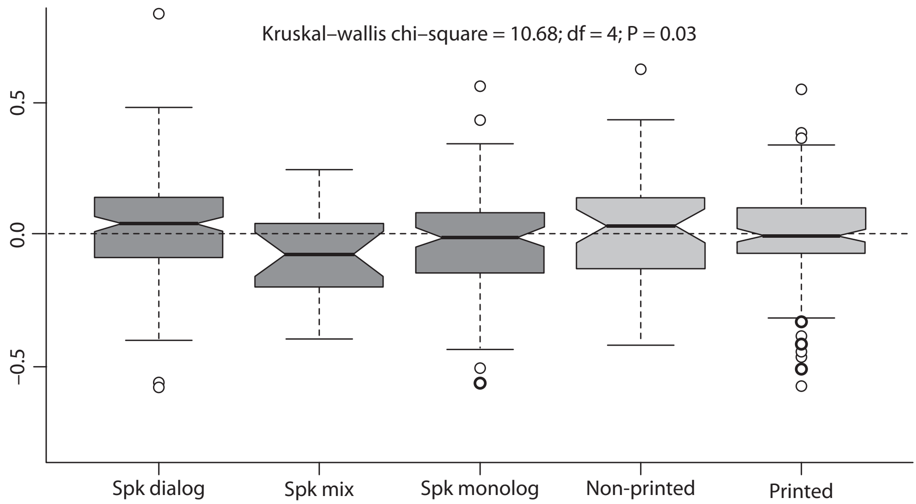


Figure 6. Percentage of improved classifications (on the y-axis) against each register (on the x-axis)

one can either exclude the file (to remove undesirable noise from the data) or one can include it nevertheless, but then the information about the (degree of) peculiarity of the data of this file/speaker is not used to make the subsequent modeling process any more precise. As mentioned above, the GLMEM approach, on the other hand, does exactly that and includes the information about particular files'/speakers' idiosyncrasies into the modeling process, which in turns makes the resulting coefficient estimates much more reliable and even allows for the interesting possibility of comparing groups of files (e.g., registers) or speakers (e.g., different sexes, different age groups, etc.). For instance, Figure 6 shows that the five registers of the ICE-GB differ significantly in terms of how much their files need to be adjusted to optimize the classification of the constructional choices: especially spoken dialog requires a lot of tweaking of the regression model whereas spoken monolog and written printed text does not.

4. Concluding remarks

So far, I have shown how – contrary to some opinions – the psycholinguistic phenomenon of syntactic priming can be studied fruitfully from a corpus-based perspective. In spite of the undisputed larger degree of noise that corpus data contain, the data provide converging evidence for several clear and significant trends attested in separate experimental studies: (i) the general strength of the repetition effect; (ii) the effect of VFormID that was found in both regression models; and (iii) the effect of VLemmaID that was found in the binary logistic regression and is compatible with the random effect of VLemmaTarget discussed in Section 3.3. In Gries (2005), I discussed how such findings can be straightforwardly related to, for example, Pickering and Branigan's account of priming based on combinatorial nodes, which is of course only one of several competing accounts. The more important point, however, is how the corpus-based data are best analyzed. Obviously, different statistical approaches are conceivable, ranging from the utterly simplistic cross-tabulation via binary logistic regression to generalized linear mixed-effects models. Just as obviously, however, the results are also quite varied.

Variationist tools like Varbrul notwithstanding, the binary logistic regression approach is probably the currently most widespread method for the kind of question studied here. However, while for the present data, the classification accuracy is significantly higher than the one obtained by chance, it is nevertheless not particularly high. More problematically, however, is the fact that the number of predictors reaching standard levels of significance – eight – is rather high and includes variables whose overall relevance to predicting one constructional choice priming is probably rather tenuous and whose effect is therefore weak (e.g., Medium, but cf. Szmrecsanyi (2006) for detailed discussion of register effects).

In some sense at least, the GLMEM approach behaves the opposite way: the number of predictors that turn out to be significant is very small – 3 – and the predictors are immediately and obviously key to priming, "obviously key" in the sense of having received support in many different studies and being integratable straightforwardly into a theoretical account based on psycholinguistic processing. At the same time, the classification accuracy is close to 90% and I have shown above, if only briefly, that this is due to the fact that the GLMEM approach includes a lot of item-specific information. This method should therefore be of interest to the population of empirical linguists – after all, it

– increases the classification accuracy;
– provides more precise coefficient estimates;
– in this case, does this with a smaller, and hence more parsimonious, set of predictors, which helps keep the analyst focus on what is really relevant once file/subject/item-specific variation is also accounted for.

It is this last part, however, that should also make this approach extremely interesting to cognitive linguists. These days, most cognitive linguists adopt a usage-based perspective in which item/exemplar-specific knowledge is essential to many aspects of language acquisition, processing, and change. With regard to language acquisition, studies by Tomasello, Lieven, Goldberg and others have illustrated that the acquisition of syntactic patterns is very much item-specific in the sense that, for instance, the acquisition of each argument structure construction is driven by one verb whose semantics match that of the construction and which, at an early stage, accounts for the vast majority of all instances of the construction (cf. Tomasello 2005 or Goldberg 2006: Ch. 4–5). With regard to processing, I have already mentioned a few examples above; recall how individual verbs' subcategorization preferences are correlated with the resolution of syntactic ambiguities and garden pathing, and Bybee and Scheibman (1999) have shown how phonological reduction processes (pointing to more automatic processing of elements having attained unit status) are correlated with individual verbs' frequency of occurrence in some syntactic pattern. With regard to language change, it is well known that, for example, grammaticalization processes are driven by the frequent co-occurrence of specific lexical items (cf. the evolution of *going to* V into a marker of futurity). In addition, we also know from, say, Dabrowska (submitted, and Dabrowska and Street 2006), that even native speakers exhibit large differences in linguistic competence. In other words, the random effects studied here target exactly and confirm the item/exemplar-specific and speaker-specific effects other work in cognitive/usage-based linguistics have uncovered.

In addition, priming studies that include more verb-specific effects than just the simple intercept adjustment I included in the GLMEM approach have even more to offer to cognitive/usage-based approaches. Snider (2008) studies prepositional dative/ditransitive as well as active/passive priming and finds that structural priming is sensitive to some of the same factors as lexical priming: high frequency structures prime less, and more similar prime and target structures prime more. Thus, these results not only support to some extent the position that syntax and lexis are not as different as many traditional (formal) approaches have assumed, but they also affect our choice of psycholinguistic models: Snider finds support for exemplar models that use clouds of exemplars represented in a feature space, but not for those that use construction-like representations. The importance of such findings for future work in cognitive linguistics can hardly be overestimated.

In sum, GLMEM approaches are not only statistically more successful, but provide also evidence converging with other contemporary cognitive-linguistic or exemplar-based approaches and are of vital importance in how they can fuel future theoretical developments. While I have so far mainly mentioned instances of converging evidence, it is worth recalling that GLMEM approaches can also offer interesting kinds of diverging evidence: since much of the item/speaker-specific variation is accounted for by the model's random effects, potentially relevant variables that are in fact more reducible to idiosyncratic variation will not be returned as significant; in other words, false positives are avoided and may speak against previous results from whatever type of (observational or experimental) approaches. Given all these advantages, I would therefore hope that this method becomes more and more widespread to improve the ways in which we as cognitive linguists look at our data.

References

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18: 355–387.

Bock, J. Kathryn. 1989. Closed-class immanence in sentence production. *Cognition* 31: 163–186.

Bock, J. Kathryn, Gary S. Dell, Franklin Chang, & Kristine H. Onishi. 2007. Persistent structural priming from language comprehension to language production. *Cognition* 104: 437–458.

Bock, J. Kathryn & Zenzi M. Griffin. 2000. The persistence of structural priming: transient activation or implicit learning. *Journal of Experimental Psychology: General* 129: 177–192.

Bock, J. Kathryn & Helga Loebell. 1990. Framing sentences. *Cognition* 35: 1–39.

Branigan, Holly P. Martin J. Pickering, & Alexandra A. Cleland. 1999. Syntactic priming in written production: evidence for rapid decay. *Psychonomic Bulletin and Review* 6: 635–640.

Branigan, Holly P., Martin J. Pickering, Simon P. Liversedge, Andrew J. Stewart, & Thomas P. Urbach. 1995. Syntactic priming: investigating the mental representation of language. *Journal of Psycholinguistic Research* 24: 489–506.

Branigan, Holly P., Martin J. Pickering, Andrew. J. Stewart, & Janet F. McLean. 2000. Syntactic priming in spoken production: linguistic and temporal interference. *Memory and Cognition* 28: 1297–1302.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, & R. Harald Baayen. 2007. Predicting the dative alternation. In G. Bouma, I. Krämer, and J. Zwarts, eds., *Cognitive Foundations of Interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Bybee, Joan & Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37: 575–596.

Chang, Franklin, Gary S. Dell, J. Kathryn Bock, & Zenzi Griffin. 2000. Structural priming as implicit learning: a comparison of models of sentence production. *Journal of Psycholinguistic Research* 29: 217–229.

Connine, Cynthia, Fernanda Ferreira, Charlie Jones, Charles Clifton, & Lyn Frazier. 1984. Verb frame preferences: Descriptive norms. *Journal of Psycholinguistic Research* 13: 307–319.

Dabrowska, Ewa. submitted. Individual differences in native language attainment: A review article.

Dabrowska, Ewa & James Street. 2006. Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Language Sciences* 28: 604–615.

Estival, Dominique. 1985. Syntactic priming of the passive in English. *Text* 5: 7–22.

Garnsey, Susan M., Neal J. Pearlmutter, Elizabeth Myers, & Melanie A. Lotocky. 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37: 58–93.

Gelman, Andrew & Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge & New York: Cambridge University Press.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Gries, Stefan Th. 2005. Syntactic priming: a corpus-based approach. *Journal of Psycholinguistic Research* 34: 365–399.

Gries, Stefan Th. under review. Frequencies, probabilities, association measures in usage-/exemplar-based linguistics: some necessary clarifications.

Gries, Stefan Th., Beate Hampe, & Doris Schönefeld. 2005. Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16: 635–676.

Gries, Stefan Th., Beate Hampe, & Doris Schönefeld. 2010. Converging evidence II: more on the association of verbs and constructions. In J. Newman & S. Rice eds., *Experimental and Empirical Methods in the Study of Conceptual Structure, Discourse, and Language*, 59–72. Stanford, CA: CSLI.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9: 97–129.

Gries, Stefan Th. & Stefanie Wulff. 2003. Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics* 3: 182–200.

Gries, Stefan Th. & Stefanie Wulff. 2009. Psycholinguistic and corpus linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics* 7: 164–187.

Hare, Mary L., Ken McRae, & Jeffrey L. Elman. 2003. Sense and structure: meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language* 48: 281–303.

Hartsuiker, Robert J., Martin J. Pickering, & Eline Veltkamp. 2004. Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science* 15: 409–414.

Jaeger, T. Florian. 2008. Categorical data analysis: away from ANOVAs transformation or not and towards Logit Mixed Models. *Journal of Memory and Language* 59: 434–446.

Jaeger, T. Florian & Neal Snider. 2007. Implicit learning and syntactic persistence: surprisal and cumulativity. In L. Wolter & J. Thorson, eds., *University of Rochester working papers in the language sciences* 3: 26–44.

Johnson, Keith. 2008. *Quantitative Methods in Linguistics*. Malden, MA: Blackwell.

Levelt, Willem J. M. & Stephanie Kelter. 1982. Surface form and memory in question answering. *Cognitive Psychology* 14: 78–106.

Loebell, Helga & J. Kathryn Bock. 2003. Structural priming across languages. *Linguistics* 41: 791–824.

Luke, Douglas A. 2004. *Multilevel Modeling*. Thousand Oaks, CA and London: Sage Publications.

Pickering, Martin J. & Branigan, Holly P. 1998. The representation of verbs: evidence from syntactic priming in language production. *Journal of Memory and Language* 39: 633–651.

Pickering, Martin J. & Holly P. Branigan. 1999. Syntactic priming in language production. *Trends in Cognitive Sciences* 3: 136–141.

Pickering, Martin J., Holly P. Branigan, Alexandra A. Cleland, & Andrew J. Stewart. 2000. Activation of syntactic information during language production. *Journal of Psycholinguistic Research* 29: 205–216.

Potter, Mary C. & Linda Lombardi. 1998. Syntactic priming in immediate recall of sentences. *Journal of Memory and Language* 38: 265–282.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org>.

Sankoff, David & Suzanne Laberge. 1978. Statistical dependence among successive occurrences of a variable in discourse. In D. Sankoff & S. Laberge, eds., *Linguistic Variation: Methods and Models,* 119–126. New York: Academic Press.

Schoonbaert, Sofie, Robert J. Hartsuiker, & Martin J. Pickering. 2007. The representation of lexical and syntactic information in bilinguals: evidence from syntactic priming. *Journal of Memory and Language* 56: 153–171.

Shin, Jeong-Ah & Kiel Christianson. 2009. Syntactic processing in Korean-English bilingual production: evidence from cross-linguistic structural priming. *Cognition* 112: 175–180.

Snider, Neal. 2008 An exemplar model of syntactic priming. Unpublished Ph.D. Dissertation, Stanford University.

Stallings, Lynne M., Maryellen C. MacDonald, & 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.

Szmrecsanyi, Benedikt. 2005. Language users as creatures of habit: a corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1: 113–150.

Szmrecsanyi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English: A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin & New York: Mouton de Gruyter.

Tomasello, Michael. 2005. *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.