9 Priming of Syntactic Alternations by Learners of English

An Analysis of Sentence-Completion and Collostructional Results

Stefan Th. Gries

General introduction

Over the last few decades, corpus-linguistic methods have become more and more of a mainstay in both theoretical and applied linguistics; in fact, frequency, dispersion, and co-occurrence data have become central notions in theoretical frameworks such as cognitive linguistics, construction grammar, or, more generally, usage-based linguistics. This cluster of theories assumes as their main building block units or *constructions*. Constructions are pairings of form and function (broadly understood) that (i) involve a lack of complete predictability of their formal and/or functional characteristics and/or (ii) a 'sufficiently high' frequency of occurrence (see Goldberg, 2006, p. 5). Constructions are considered to vary in size/complexity from simple morphemes up to fairly abstract sentence-level or argument structure constructions.

For quite some time, much work in construction grammar was somewhat narrow both in terms of methodology and scope: With regard to the former, construction grammar was as largely introspective as the generative theories of language it argued against. With regard to the latter, construction grammar has focussed much more on the constructicon—the construction-grammar equivalent to the lexicon—of the native speaker as opposed to that of non-native speakers such as speakers of (mostly) English as a second or a foreign language.

By now, times have changed and construction grammar has come to rely much more on corpus data to obtain frequencies of (co-)occurrence of constructions that are argued to be correlated with L1-acquisition, language processing, use, and change, but also the language systems of L2/FL-speakers.

For a theory that essentially claims that (i) linguistic knowledge *is* knowledge of constructions in the described sense and that (ii) linguistic structure and representation emerges from language use, one obvious question or area of research is of course the degree to which non-native speakers' constructicons are comparable to that of native speakers: Do non-native speakers build up their constructicon(s) in a way similar to native speakers? To what degree do non-native speakers' constructicons differ from those of native speakers, given the limited amount of input and already existing but different L1 system of the former? Are the differences quantitative tendencies or (bigger) qualitative differences? These are the kinds of questions the present chapter is concerned with. To that end, this chapter uses a combination of corpus-linguistic and experimental (sentence-completion) data on the so-called dative alternation between ditransitives and, here, *to*-datives exemplified in (1a) and (1b) respectively.

(1) a. Picard gave [$_{\text{NP Rec}}$ the Borg] [$_{\text{NP Pat}}$ his phaser].
 b. Picard gave [$_{\text{NP Pat}}$ his phaser] to [$_{\text{NP Rec}}$ the Borg].

Specifically, the questions to be addressed here are the following:

- to what degree, if any, do the experimentally-obtained constructional choices of non-native speakers of English—here, advanced German learners of English—exhibit the kind of structural priming effects that have been observed in more than 30 years of priming studies?
- to the degree that non-native speakers of English exhibit such priming effects, what determines the size and their nature? Is priming, or are constructional choices, affected by characteristics of the prime (which is further away from the constructional choice) or the target (right before the constructional choice?
- to what degree, if any, are these experimentally obtained constructional choices of German learners of English correlated with the probabilistic co-occurrence distributions of verbs and constructions in native-speaker corpus data?

Previous work on these questions

To a limited extent, these questions for this alternation have been studied before, in a sentence-completion experiment involving learners of English reported on in Gries and Wulff (2005), which was a replication of a series of experiments done by Pickering and Branigan (1998) with native speakers of English. However, their analysis was undertaken before the more widespread acceptance of the more advanced mixed-effects modeling that is now typically used for such data. Instead, Gries and Wulff (2005) conflated all their prime and target frequencies from different conditions, speakers, and stimuli into one prime × target frequency table, for which they computed inferential statistics and an effect size. Their results are shown here as Table 9.1, for which we could compute an odds ratio of 2.57 (with a 95% confidence interval of (1.85, 3.58)), which would be how much a certain prime construction affects—here, boosts—the odds of seeing the same target construction.

In addition, they found an overall correlation between the verbs' constructional preferences in the learners' sentence completions on the one

Table 9.1 Construction Frequencies Obtained in the Priming Experiment (Gries & Wulff's (2005) Table 1)

	Target: Prepositional Dative	*Target: Ditransitive*	*Totals*
Prime: prepositional dative	186	146	332
Prime: ditransitive	106	214	320
Totals	292	360	652

hand and in native-speaker corpus data on the other; both results they interpreted as evidence for some sort of constructional knowledge/representation on the part of the learners (because only something that is mentally represented in some way can be primed) and one that appears to be heavily usage-based (because it exhibited very similar probabilistic preferences to native speaker usage of constructions).

While their results 'made sense', were encouraging, and received some support from a later priming study on *to*/*ing*-complementation (Gries & Wulff, 2009), the resolution of their analysis is relatively coarse and the field in general has undergone quite some development regarding quantitative sophistication in the meantime. It is useful to determine whether the findings are in fact accurate in the sense that both their general priming effects and their specific correlation between learner and native-speaker choices remain even when a much more sophisticated statistical analysis is applied to the totality of their data; this is because it is not rare for mixed-effects regression modeling to reveal that much of the variability in the data that a fixed-effects-only analysis, or one that conflated much of the data in the way Gries and Wulff (2005) did, is in fact due to speaker-specific and/or stimulus-specific effects that one is usually not interested in and would control for better given how the field has evolved since then.

The purpose of this chapter, therefore, is to tackle the preceding three research questions utilizing all the information that Gries and Wulff (2005) did not incorporate into their evaluation. At the same time and in keeping with this volume's thematic focus, I will also use this analysis to explore the convergence, or divergence, of the experimental sentence-completion data and the observational corpus data, plus I will comment on how the present study shows how much corpus data and the logic underlying them can contribute even to areas of research that, until very recently, were firmly in experimental-psycholinguistics territory.

Methods

This section outlines how the experimental and observational data analyzed here were obtained and how they were then analyzed statistically.

Experimental Data

In this section, I will discuss the ways in which the experimental data were obtained. The experiment conducted was a replication of Pickering and Branigan's (1998) series of priming experiments with native speakers of English using their stimuli and sentence-completion paradigm with advanced German learners of English (mean number of years of English teaching: 11.1 years, interquartile range: 2.6 years); while no individual proficiency scores were obtained, it is very likely that nearly all of these learners would be considered as B2 or C1 speakers in the CEFR. Sixty-four subjects received a questionnaire with 32 sentence fragments in a random order and were instructed to complete each fragment such that the result constituted a grammatically correct sentence. Sixteen of the 32 fragments were filler items unrelated in lexical and constructional materials to the dative alternation (e.g., intransitive verbs, NP fragments ending with a relative pronoun, complete clauses to which an adverbial or a second clause could be added, etc.). The relevant 16 experimental items per questionnaire consisted of eight prime-and-target sentence-fragment pairs. Specifically, half the primes were sentence fragments that were designed to bias the sentence completion in the direction of a ditransitive (as in (2a) whereas the other half of the primes were designed to bias the sentence completion in the direction of a prepositional dative (as in (2b)):

(2) a. The racing driver showed the helpful mechanic ______________.
b. The racing driver showed the torn overall ______________.

The target fragments after the primes did not contain a postverbal NP so the subjects had to decide on a syntactic structure to complete the fragment with. Primes and targets also differed with regard to the tense, aspect, and number of the prime and target verbs; in Pickering and Branigan's study, each of these features was tested in a separate experiment to determine whether an overall expected priming effect was moderated by these morphological differences (see their appendix on pp. 647–650 for a complete list of the stimuli of all five experiments). The $64 \times 32 = 2048$ responses obtained from the subjects were then trimmed down to only those target completions that were either ditransitive and prepositional datives: this variable TARGET_COMPL_CX was the binary dependent variable in the subsequent regression analysis.

Each completion was then also annotated with regard to the following variables:

- PRIME_COMPL_CX: the construction the subjects chose to complete the prime sentence fragment with, *ditransitive* vs. *prepositional*

dative vs. *other*; this is the only variable Gries & Wulff (2005) used as a 'predictor' for the learners' target completions, whereas the present analysis will of course incorporate this one, but also all the following ones;
- PRIME_STIM_TENSE: the tense of the verb of the prime sentence fragment: *present* (e.g. *The lifeguard shows the life belt* ___) vs. *past* (e.g. *The lifeguard showed the life belt* ___);
- PRIME_STIM_ASPECT: the aspect of the verb of the prime sentence fragment: *perfective* (e.g. *The lifeguard showed the life belt* ___) vs. *imperfective* (e.g. *The lifeguard was showing the life belt* ___);
- PRIME_STIM_NUMBER: the number of the verb of the prime sentence fragment: *singular* (e.g. *The lifeguard shows the life belt* ___) vs. *plural*; (e.g. *The lifeguards show the life belt* ___);
- PRIME_STIM_VPREF: the construction towards which the verb in the prime sentence biases completion (as exemplified in (2)), *ditransitive* vs. *prepositional dative*;
- PRIME_STIM_V: the verb in the prime sentence fragment sentence that the subjects were asked to complete, *give* vs. *hand* vs. *lend* vs. *loan* vs. *post* vs. *sell* vs. *send* vs. *show*;
- TARGET_STIM_V: the verb in the target sentence fragment the subjects were asked to complete, *give* vs. *hand* vs. *lend* vs. *loan* vs. *post* vs. *send* vs. *show*;
- ITEM_NO_WOUTFILLER: the number of the current target stimulus per subject, a number ranging from 1 to 16 included as a control variable to statistically control for learning effects over the course of the experiment.

A brief comment is in order why the variables PRIME_STIM_V and TARGET_STIM_V are treated as fixed effects here when one might feel they could (also) be random ones. At this point, there is still sometimes some disagreement about what the necessary and sufficient conditions are for some variable to be a random or a fixed effect (see Gelman & Hill, 2007 for discussion). One definition is to consider something a random effect if the variable levels observed in a data set do not exhaust all levels one would observe in the sampled population; this would include TARGET_STIM_V since, obviously, there are more dative-alternating verbs than the ones tested here.

On the other hand, a random effect also presupposes that the levels of the relevant variable are a *random* sample of the population, something that is already hard to justify, though routinely done, but much harder to justify for the verbs used in the construction of the stimuli (just consider the fact that the relevant verbs are all fairly frequent straightforward transfer verbs). In addition, random effects are most useful when there is a larger number of levels, e.g. a dozen or more each attested a 'decent' number of times, and the former at least is not the case here either. Therefore, the variables PRIME_STIM_V and TARGET_STIM_V were incorporated here as fixed effects. However, the analysis did include random effects as well, which had to do with the prime and target sentence fragments as well as with the subjects filling out the questionnaires; more details on the random-effects structure and its exploration in the statistical analysis are provided next.

Corpus Data

The corpus data, to which the experimental data will be compared, consist of the results Gries and Stefanowitsch (2004) published for the dative alternation. Using the method of distinctive collexeme analysis, they computed for each verb participating in the dative alternation within the British Component of the International Corpus of English a score that summarizes which of the two constructions of the dative alternation the verb is more strongly attracted to. The association measure they used was the (negative) $\log_{10}$ p-value of the Fisher-Yates exact test (p_{FYE}) computed for 2×2 co-occurrence tables for each verb's occurrence in both constructions; see Table 9.2 for the relevant frequencies for *give* mentioned in Gries and Stefanowitsch (2004, p. 106), with the relevant computations shown in (3); the value of 119.7361 would be the distinctive collexeme strength of give toward the ditransitive (as opposed to the prepositional dative).

(3) a. –log10(sum(dhyper(461:607, 607, 2347, 1035))) # = 119.7361
    b. –log10(sum(dhyper(0: 146, 607, 2347, 1919))) # = 119.7361

Since this approach conflates association and frequency—the *p*-value would change considerably if all numbers in Table 9.2 were only 10% the size of what they are now—it may be prudent to also consider a measure of distinctive collexeme strength that quantifies association separately from frequency. One such measure that is particularly useful in the present connection is the uni-directional measure $\Delta P_{\text{construction|verb}}$, which is quantified as shown in (4).

Table 9.2 Observed Dative-alternation Construction Frequencies With *Give* in the ICE-GB (Data from Gries & Stefanowitsch 2004, Table 2)

	Construction: ditransitive	*Construction: prep. dative*	*Totals*
give	461	146	607
other verbs	574	1773	2347
Totals	1035	1919	2954

(4) a. $\Delta P_{\text{prep.dative}|give} = 146 / 607 - 1773 / 2347 = -0.5149$
b. $\Delta P_{\text{ditransitive}|give} = 461 / 607 - 574 / 2347 = 0.5149$

This measure is useful here, first, because it would not change the same way p_{FYE} does when, say, all numbers in Table 9.2 were divided by 10. Second, it is useful in how it falls between −1 (indicating strong repulsion) and +1 (indicating strong attraction). Third and to some degree most importantly, it is useful because it is a directional measure, here in its version 'from the verb to the construction' and thus, essentially asking the question 'how much does seeing the verb make you choose a construction?', which corresponds well to the experimental setup of the sentence-completion task. Here, $\Delta P_{\text{prep. dative}|give}$ is high (in the sense of 'far away from 0') and negative, reflecting that *give* is a good 'cue' for not using the prepositional dative, but the ditransitive, construction.

For the exploration of whether any experimentally-obtained constructional preferences of the verbs in the target sentence fragments are correlated with the corpus-based constructional preferences of the same verbs, verb-specific findings from the regression model on the experimental data will be correlated with both measures of distinctive collexeme strength—p_{FYE} and ΔP.

Statistical Evaluation

As mentioned earlier, in Gries and Wulff (2005), the data were analyzed in a relatively simple way given that multifactorial studies of syntactic variation outside of sociolinguistics were not yet widely established and mixed-effects modeling was fairly unknown at the time. Essentially, their analysis, while done as a chi-squared test, can be paraphrased as a binary logistic regression model with TARGET_COMPL_CX as the dependent variable and PRIME_COMPL_CX as the only predictor.

For this chapter, their data were re-analyzed in what is now a much more appropriate way, with a generalized linear mixed-effects model. Specifically, here, the following strategy was adopted. First, I explored the random-effects structure of the regression model. Because of the relatively small sample size, a maximal random-effects structure in the sense of Barr et al. (2013) was unfortunately out of the question, but given the experimental design, there were nevertheless multiple random effects to be explored:

- varying intercepts for individual subjects nested into varying intercepts for the experiments in which subjects participated;
- varying intercepts for the stimulus that served as the prime;
- varying intercepts for the stimulus that served as the target.

The random-effects structure to be sought for the subsequent modeling process was one that involved only random effects that accounted for some meaningful variability within the model.

Second, after a useful random-effects structure was identified, I attempted to determine the fixed-effects structure of the minimal adequate model with the previously determined random-effects structure. This was done by starting from the null model with the preceding random-effects and then checking which change of the model—adding or deleting any predictor—would improve the model's *AIC* most while at the same time checking every model selection step for multicollinearity (using the models' variance inflation factors (*VIF*s)) and for overdispersion; the maximal model providing the predictors to be considered for addition involved all listed main effects and all their pairwise interactions.

The final model arrived at in that way was evaluated with regard to its significance values (given that it was *AIC*-values that determined the model selection), and its classification power (using R^2-values as well as classification accuracy and, more importantly, its *C*-score). Finally, in order to be able to interpret the results, both the fixed and the random effects of the final model were explored, the former by visualizing predicted probabilities for each and every effect, the latter most importantly by computing the correlations between the predicted probabilities for prime and/or target stimulus verbs (depending on any of these effects making it into the final model) with the corpus-based preferences of the same verbs in the corpus data of the native speakers.

Results

In this section, I summarize the results of the statistical model selection process for the experimental data.

Overall Results

The planned analysis as discussed earlier and initial exploration of the data led to some trimming of the data. First, all cases where the prime and/or the target fragment completion were neither a ditransitive nor a prepositional dative were discarded. Second, all cases with the prime stimulus verbs *offer* and *throw* were discarded because they amounted to only six and two cases respectively and would therefore have caused problems for the regression modeling. Third, after that step, all cases with the target stimulus verbs *sell* were discarded because they amounted to only five cases.

The overall results of the then-following regression of the subjects' sentence completions were somewhat mixed. Several null models without any fixed-effects predictors led to a random-effects structure with varying intercepts for each subject, each prime sentence fragment, and each target

Table 9.3 Fixed Effects in the Final Regression Model (LR-tests for Deletion)

	Chi-squared	*df*	*p*
PRIME_COMPL_CX	39.09	1	<0.001
TARGET_STIM_V	41.35	6	<0.001
ITEM_NO_WOUTFILLER	2.45	1	0.117
TARGET_STIM_V : ITEM_NO_WOUTFILLER	14.89	6	0.021

sentence fragment. In this final null model, from which the selection of the fixed-effects structure was begun, the experimental subjects accounted for by far the largest amount of variability (variance = 1.9125), whereas the prime and target stimuli accounted for much less (variances = 0.1903 and 0.3164 respectively).

The subsequent exploration of the fixed-effects structure led to a final model with the predictors listed in Table 9.3; the model exhibits some collinearity, but not too much (all *VIF*s < 8.6) and no overdispersion ($p > 0.9$).

The model's R^2-values indicate only a moderate degree of correlation, $R^2_{\text{marginal}} = 0.192$ and $R^2_{\text{conditional}} = 0.523$, indicating that much of the structure in the data is contributed by the between-subjects variability rather than the fixed effects usually of interest. On the other hand, when that between-subjects variability is taken into account, the model has a decent amount of classification accuracy (81.1%), which is highly significantly better than chance/baseline, and a good *C*-score (0.89). The following two sections will discuss the fixed- and random-effects structures respectively.

The Fixed-effects Structure

The first fixed effect to be discussed is PRIME_COMPL_CX, i.e. the construction that the subjects chose to complete the prime sentence fragments with. Its effect is represented in Figure 9.1: The x-axis represents the two levels of the predictor, the y-axis represents the range of predicted probabilities of prepositional datives in the target sentence fragment, the black points and intervals are the predicted probabilities (of 0.4 and 0.73 for ditransitive and prepositional dative primes respectively) and their 95% confidence intervals, and the plotted points are the individual predicted probabilities for each level of the predictor, "with light grey and dark grey representing correct and incorrect classifications".

The effect is relatively straightforward: When the subjects completed the prime with a prepositional dative, then they were also much more likely to complete the target with a prepositional dative and vice versa; the odds ratio for that effect is 4 (95% confidence interval from 50 bootstrapped models: (2.4, 6.18)). This is clear evidence for a self-priming effect and the fact that this predictor does not interact with any other one

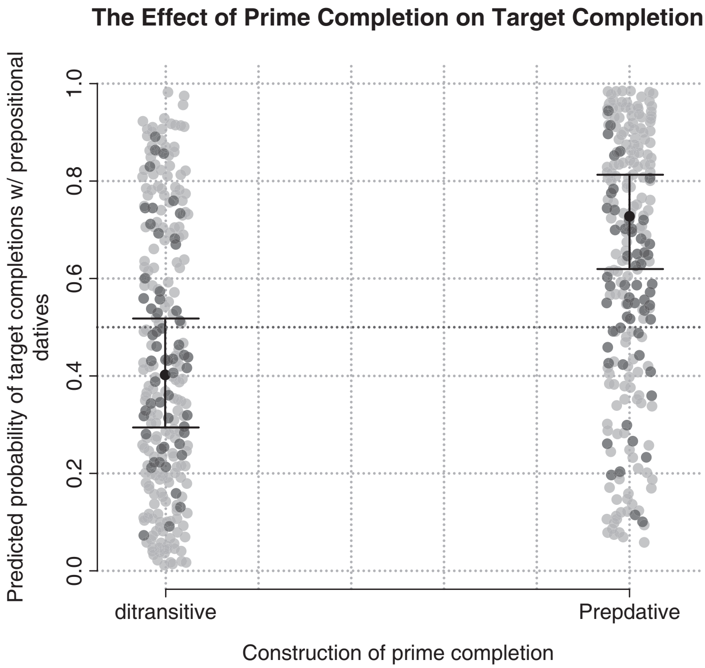

Figure 9.1 The Effect of PRIME_COMPL_CX on the Predicted Probability of Completing the Target Sentence Fragment With a Prepositional Dative

also reveals that this self-priming effect is independent of which verbs the prime and the target contained as well as what the prime verb's preference was (an important point to which we will return later in the section Discussion and Concluding Remarks when I discuss the predictor surprisal, a proxy for processing difficulty).

This being said, the data also show that the verb in the target sentence fragment had a significant impact, too, although, as we will see presently, one that is moderated by within-experiment learning effects. This main effect of TARGET_STIM_V is shown in an analogous plot in Figure 9.2.

Several aspects of this predictor are noteworthy. First, it is clear that *give*, the prototypical ditransitive literal-transfer verb, has, as might have been expected from rather advanced learners of English, the highest *dis*-preference for prepositional datives (predicted probability of prepositional datives = 0.3) and, thus, the highest preference for ditransitives. The only other verb that comes close to *give* in that regard is *show* (predicted probability of prepositional datives = 0.39). On the other hand, the verbs that are most strongly associated with a prepositional dative completion are *post* and *send*, closely followed by *loan* (predicted probability of prepositional

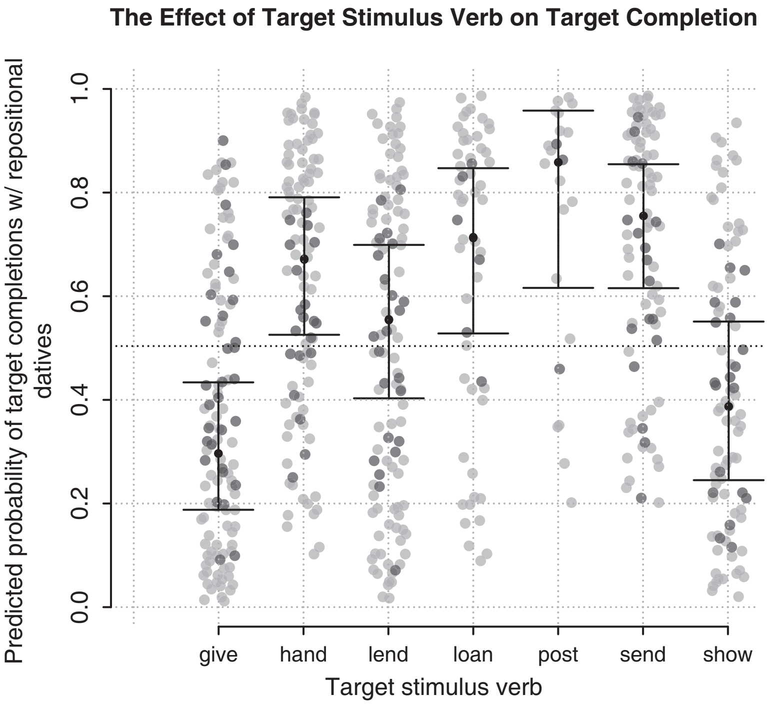

Figure 9.2 The Effect of TARGET_STIM_V on the Predicted Probability of Completing the Target Sentence Fragment With a Prepositional Dative

datives = 0.86, 0.75, and 0.55 respectively). This is interesting in how it corroborates at least tendentially iconicity analyses of the dative alternation such as Thompson and Koide (1987): The verbs that prefer the ditransitive most strongly are *give* and *show*, which typically involve a small spatial distance between the agent and the recipient, whereas the verbs that prefer the prepositional dative most strongly are *post* and *send*, which typically involve a much larger distance between agent and recipient, which has been argued to correlate with the larger distance between the linguistic signs denoting them and the path-denoting preposition *to*.

In spite of the preceding, it needs to be borne in mind that the prior main effect needs to be taken with a grain of salt because it participates in an interaction with ITEM_NO_WOUTFILLER, the variable included to control for within-experiment learning effects. This effect is visually represented in Figure 9.3, with ITEM_NO_WOUTFILLER on the x-axis, predicted probabilities of the prepositional dative on the y-axis (as usual), and the different verbs in the target sentence fragment represented with differently-colored lines.

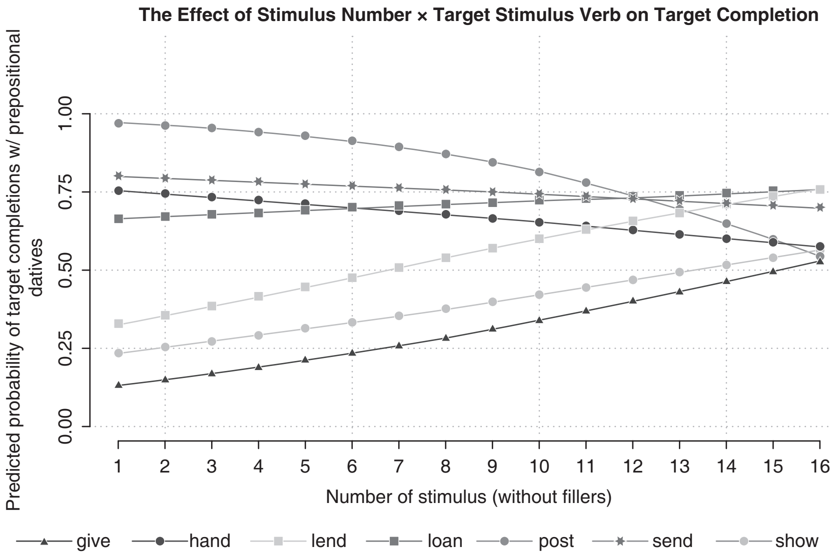

Figure 9.3 The Effect of ITEM_NO_WOUTFILLER : TARGET_STIM_V on the Predicted Probability of Completing the Target Sentence Fragment With a Prepositional Dative

This interaction is interesting both in terms of methodology and in terms of linguistic interpretation. The graph clearly shows that, especially with the maybe more malleable language systems of learners (as opposed to those of native speakers), even just the exposures over the course of an experiment can trigger learning effects. This confirms results from Gries and Wulff (2009) or Doğruöz and Gries (2012), who both found that the subjects in their experiments changed their tendencies to react to the experimental stimuli over the course of as few as eight exposures to experimental stimuli; at the same time, this is exactly what Jaeger (2010, p. 53) discusses well in his argument of how, in experiments with balanced designs, subjects "are put into a situation where words and syntactic structures (co-)occur in (uniform) distributions that do not match participants' expectations based on previous experience with naturally distributed data," a finding that also compatible with how priming can be cumulative (see Jaeger & Snider, 2008). Thus, one should in fact always expect a certain clash between learners' previous distributional experience and the unnatural distributions of nicely balanced experiments and control for it at least statistically after the fact.

In the present case, having included this statistical control allows us to see how, over the course of the experiment, the main effect discussed earlier changes: We saw that *give* and *show* are most strongly associated with the ditransitive whereas *send* and *post* are most strongly associated with the prepositional dative, findings that are perfectly compatible with previous work. Here, however, we see that these effects are only obtained well in about the first half of the experiment because, as we get to stimulus numbers 10, 12, etc. these four verbs' constructional preferences have been diluted, so to speak.

The Random-effects Structure and its Relation to the Corpus Data

One interesting side effect of the kind of statistical analysis performed here is that it allows for a relatively straightforward comparison of the verbs' constructional preferences arrived at on the basis of corpus-linguistic data and the verbs' constructional preferences obtained from the subjects' sentence completions—either as varying intercepts/slopes or as, here, coefficients of fixed effects in a regression model.

In the present example, we can look up the constructional preferences of the stimulus verbs in the native speaker corpus data—the ICE-GB—from Gries and Stefanowitsch (2004) and in fact compute additional measures that they did not report in that chapter: They report, as most (distinctive) collexeme analyses do, the (often $\log_{10}$-transformed) *p*-values of Fisher-Yates exact tests, a bidirectional association measure first proposed by Pedersen (1996) that, as a significance test, is affected by both effect size/association strength and sample size. However, as previously discussed, one can also compute alternative measures such as a version of ΔP, a directional measure of association that can distinguish between cases where a verb attracts a construction as opposed to a construction attracting a verb and, as a difference of percentages, is not affected by sample size. Table 9.4 summarizes these corpus-based measures on the basis of the Gries & Stefanowitsch (2004) data; note that the $\log_{10}$ged p_{FYE}-values were set to positive and negative when the verb was attracted to the prepositional dative and the ditransitive respectively to make sure all scales have the same orientation.

The rank correlations between the three measures are fairly high: Spearman's *rho* for the correlation between $-\log_{10} p_{FYE}$-values and the predicted probabilities following from the main effect of TARGET_STIM_V is 0.714 (but not significant: $p_{1\text{-tailed}} = 0.068$). However and more importantly, the same coefficient for the correlation between ΔP

Table 9.4 Collexeme Strengths of Verbs From Gries and Stefanowitsch's (2004) ICE-GB Data Together With Predicted Probabilities for Target Fragment Verbs in Figure 3.2

	give	*hand*	*lend*	*post*	*sell*	*send*	*show*
$-\log_{10} p_{FYE}$	−119.736	1.197	0.222	0.937	1.857	−0.396	−11.08
$\Delta P_{prepdativelv}$	−0.515	0.159	0	0.351	0.285	−0.012	−0.424
pred. probs.	0.296	0.671	0.555	0.857	NA	0.754	0.388

Note: Oriented such that positive values indicate a preference for prepositional datives and the predicted probabilities following from the main effect of TARGET_STIM_V is 0.829 (which is significant: $p_{1\text{-tailed}} = 0.0292$), and Spearman's *rho* for the correlation between ΔP and the observed probabilities in the completion results is even higher with 0.9 (also significant with $p_{1\text{-tailed}} = 0.007$). Thus, all results indicate a rather high degree of compatibility between the native speaker corpus data and the probabilistic preferences of the learners as manifested in their sentence completions. This strongly suggests that the representation of the constructions participating in the dative alternation with *to* and the verbs they go with in advanced learners' probabilistic language system is very similar to that of native speakers. However, note that this tendency is only significant with the directional association measure (i) that does not include frequencies but only the effect size and (ii) whose orientation is compatible with the experimental design, which went from verb to construction.

There is one final kind of exploration that can be insightful, which I will mention here only for the sake of completeness (and because few studies do it): The exploration of how the varying intercepts per speaker affect the speakers' predicted responses/completions (and, correspondingly, make the model better). Figure 9.4 represents those results: The *x*-axis

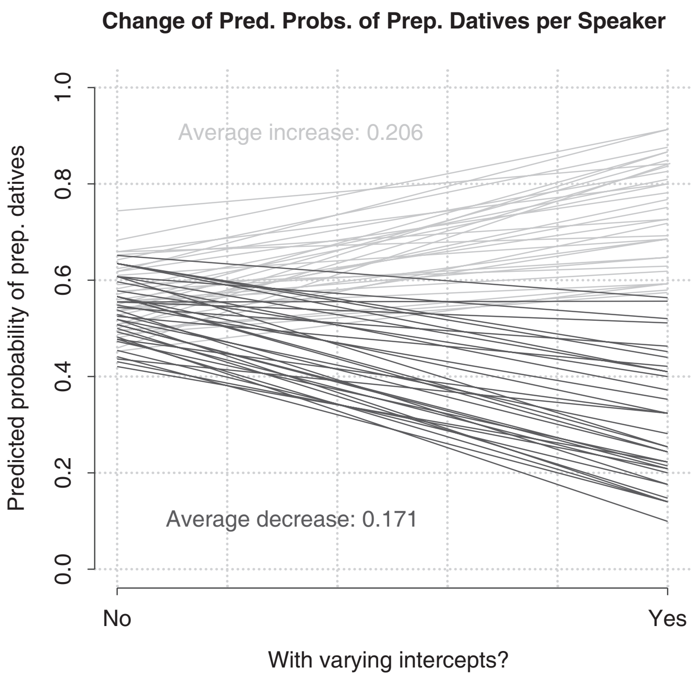

Figure 9.4 The Effect That Speaker-Specific Intercepts Have on the Average Predicated Probability of Completing the Target Sentence Fragment With a Prepositional Dative

distinguishes between predictions without speaker-specific intercepts (left) and with speaker-specific intercepts (right), the *y*-axis, as always, represents predicted probabilities of prepositional datives, and each line represents, for one speaker, how the average prediction for that speaker changes depending on whether his overall preference for a construction is accounted for: "If the line is light grey", a speaker's preference for the prepositional dative *in*creased the prepositional dative's probability by on average 20.6%—"if the line is dark grey", a speaker's preference for the ditransitive *de*creased the prepositional dative's probability by on average 17.1%. Such summary results or the exact speaker-specific intercepts could be correlated with other speaker-specific information such as foreign language proficiency scores, but also more general speaker-specific factors such as personality, aptitude, and motivation factors to help account for individual variation, a growing topic in second language acquisition and learner corpus research.

Discussion and Concluding Remarks

Returning to the Research Questions

As outlined earlier, this chapter is concerned with three questions, which will be recapitulated and addressed in this section. The first question was whether constructional choices by advanced non-native speakers of English exhibit the same kind of structural priming effects as native speakers have done in hundreds of studies over the last 30 years. For the most part, this question can be answered in the affirmative: The learners exhibit significant production-to-production priming effects such that they prefer to complete the target sentence fragment in the same way as they completed the prime fragment even when many other factors including speaker- and stimuli-specific idiosyncrasies are controlled for. In addition to supporting the role of priming, these results are also compatible with Gries and Wulff's (2009) experimental results on *to* vs. *ing* complementation as exemplified in (5).

(4) a. Riker tried to rescue Picard from the Borg.
    b. Riker tried rescuing Picard from the Borg.

In a sentence-completion task similar to the one reported on here, they, too, find that the completion of the prime has a significant effect on the completion of the target (Gries & Wulff, 2009, p. 176). In addition and even more encouragingly from a triangulation/converging evidence perspective,

- as mentioned previously, the odds ratio for the overall completely conflated priming effect in Gries & Wulff (2005) was 2.57;

- the probably most influential priming study, Bock (1986), reports priming results for the dative alternation that correspond to an odds ratio of 3.3.

Crucially, both of these fall squarely within the 95% confidence interval of the corresponding odds ratio (of 4, see earlier section The Fixed Effects Structure) in this chapter's regression model, providing for excellent convergence of results across very different experimental designs and statistical analyses.

The second question was what the main determinants of, here, learner priming would be. In the present study, both characteristics of the prime and of the target sentence fragments co-determined subjects' completions; however, a direct effect size comparison is hard to make because of how the target stimulus verb interacts with the control variable of the stimulus number.

The third and final question was whether the learners' completion choices were correlated with the (prime and/or target) verbs' preferences. The results show that that is clearly the case: While there appear to be learning effects over the course of the experiment, there is also a clear effect of which verb the target completion is written for, and those probabilistic preferences are strongly correlated with the same verbs' constructional preferences in native-speaker data regardless of whether they are measured bidirectionally and with frequency built into the measure (p_{FYE}) or unidirectionally and without frequency ($\Delta P_{\mathrm{construction|verb}}$). These results, too, are compatible with Gries and Wulff (2009), in whose regression model verb-specific preferences had a big effect on *to* vs. *ing* completions. This result is also not very different from the finding reported in Gries and Wulff (2005), but it is nonetheless important to point out that this convergence is still far from trivial because it has been arrived at here in a way that is much more robust, comprehensive, and more the state-of-the-art than in their study: Given how standards evolved since 2005, the present approach includes

- a *multifactorial analysis* in which the many causes that could have given rise to their results are distinguished and contrasted with each other within one and the same modeling approach; this for instance made it possible to see what seems to have been responsible for the priming effect they obtained in their coarser analysis.
- a *random-effects structure* that, within the limits imposed by the sample size, took stimuli and speaker variability into consideration, making it harder for individual-stimulus/-speaker results to skew the overall findings in a certain direction.
- a statistical control for *learning over the course of the experiment* (not unlike, but more advanced than, the related later study of Gries & Wulff, 2009).

- a *better comparison* between the (better) experimental results and the corpus data by adding a measure that, unlike most previous analyses, is not bidirectional and does not also incorporate frequencies ($\Delta P_{\text{construction|verb}}$).

It was therefore not obvious at all that (i) the current more appropriate reanalysis of their data would yield a significant regression model for the experimental data (something they did not perform) and, therefore and more importantly, that (ii) the verb-specific results of that regression model would still correlate with the differently measured corpus-based preferences especially given that the former are now just one of many predictors in a multifactorial model.

On Methodological Triangulation

What are the results' implications regarding methodological triangulation? In some sense at least, the results could hardly underscore the relevance of methodological triangulation and converging evidence more. It was already discussed previously how the present results, in particular the learning effect over time, underscore the potential risks of balanced experimental designs that expose speakers—native or otherwise—to very unrepresentative input distributions: As was shown earlier, learning effects can be observed even over a relatively small number of experimental stimuli that lead to verb-specifically different sentence-completion patterns over experiment-time. However, there are even more issues that might give rise to concerns. This is because the input distribution to the subjects is not only unrepresentative when it comes to the coupling of the verbs in the prime sentence fragments and their likely constructional continuations, which is much closer to 50:50 than how these verbs are actually used in naturally-occurring data—the input distributions are also unrepresentative in ways that relate to two other variables that moderate priming: prime-target similarity (see Snider, 2009) and surprisal (Jaeger & Snider, 2008).

As for the former, Snider shows that priming is stronger when prime and target are similar to each other (see also Scheepers, Raffay, & Myachykov, 2017). While this is not completely surprising and in fact one reason why Pickering and Branigan's (1998) included prime-target verb pairings that differed or did not differ with regard to tense, aspect, and number, Snider's corpus study is much more comprehensive in how he computes a multidimensional similarity metric (the Gower metric) that quantifies the overall similarity between every prime and target. Crucially, even in a regression model that includes many other factors to predict the dative alternation, he still finds a significant interaction between the construction of the prime and the similarity of the prime to the target such that higher similarity between prime and target boosts priming strength. Obviously, the usually highly controlled sentences characteristic of experiments whose similarity serves to control noise variables will introduce an unrepresentatively high degree of similarity into the subjects' input that bears little resemblance to usually much more varied actual discourse.

As for the latter, the notion of surprisal has now been included in a variety of studies. Following Hale (2001, p. 4), who in turn refers back to work as early as Attneave (1959), surprisal can be defined as "the combined difficulty of disconfirming all disconfirmable structures at a given word" or, more mathematically, $\log_2 ({}^{1}/_{p(\text{word i} \mid \text{word i-1}, \ldots, \text{context})})$ or $-\log_2 p(\text{word}_i \mid \text{word}_{i\text{-}1}, \ldots, \text{context})$; thus, surprisal is a heuristic measure of processing difficulty. However, surprisal need not only be considered as a function of transitional probabilities—it can occur with any kind of conditional probabilities, including the question of whether a certain verb occurs with the construction it is expected with. For example, Jaeger and Snider (2008) show that surprisal manifests itself in priming as the degree to which a construction primes more if it is less expected given its (lexical) context. This makes their notion of surprisal very similar to distinctive collexeme strength, just not of the target verb as explored earlier but of the prime verb. This is relevant here because, maybe in part because of the skewed distribution of the input the subjects received, this study did not find any such effect. Surprisal was operationalized as the potential interaction of PRIME_STIM_V and PRIME_STIM_VPREF: is the prime verb provided in a sentence fragment favoring the construction the prime verb prefers (low surprisal) or not (high surprisal)? However, no effects involving PRIME_STIM_V or PRIME_STIM_V : PRIME_STIM_VPREF were in the final regression model, nor did forcing them in, or forcing in an interaction of these with the subjects' completion of the prime, result in any improvement of the regression model—in fact, even the most lenient/benevolent approach, fitting a separate logistic regression model with only surprisal as a predictor, i.e. a situation where surprisal might 'soak up' all sorts of variability in the data that, normally, other predictors would control for, did not lead to surprisal playing a significant role for predicting speakers' completion or to surprisal correlating well with the predicted probabilities of prepositional datives per verb (Spearman's *rho* = –0.486). While this is not incontrovertible additional *proof* regarding the role of the skewness of the input, it certainly is a suggestive addition to the other results and arguments that point to the potential problematic effects of balanced experimental designs, and it is unlikely that this lack of an effect can be explained by simply pointing to the fact that the speakers tested are not native speakers, given the otherwise very close correspondence between their behavior in the experiment and native-speaker corpus data.

Given all of the preceding, the importance of methodological triangulation should be self-evident. While priming research has been predominantly experimental and even been somewhat critical of the first more corpus-based studies that appeared 10–15 years ago (see Branigan, Pickering, Liversedge, Stewart, & Urbach, 1995, p. 492; Pickering & Branigan, 1999, p. 136), I think there is now more widespread recognition of what corpus-based psycholinguistics and/or computational psycholinguistics have to offer: Data and analyses that are messier than much more traditional psycholinguistic work has had to deal with, but that can offer a degree of ecological validity that even careful experimentation cannot always guarantee.

That being said, the degree to which observational corpus data can provide a complementary contribution to experimental data is greatly dependent on how much the inherent noise/messiness of observational data can be controlled for statistically and after the fact. Corpus-based/computational psycholinguistics has made great strides with regard to the methodological challenges that observational data pose but, as always, much more needs to happen in order for researchers to be able to strike the best balance possible between clean control and messy ecological validity. It seems to me as if linguists have a long tradition of looking to psychology for methodological tools and inspiration, which of course has a firm foundation of many decades of refining the statistical analysis of experimental data. My personal impression is, however, that for the best kinds of analysis of *observational* data, linguists are well advised to consider (also) other fields, in particular ecology, where observational researchers have to deal with quite similar and similarly noisy data in habitats and, therefore, have had to develop solutions to problems observational/corpus researchers in linguistics are now beginning to struggle with. As so often, therefore, I hope that a growing degree of interdisciplinariness will help the field to develop the right kinds of methodological tools, which in turn would facilitate much needed methodological triangulation, of which the present chapter was just a small example.

References

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods and results*. New York: Holt, Rinehart and Winston.

Barr, D., Roger, L., Christoph, S., & Harry, J. T. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387.

Branigan, H. P., Pickering, M. P., Liversedge, S. P., Stewart, A. J., & Urbach, T. P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research*, *24*(6), 489–506.

Doğruöz, A. S., & Stefan, Th. G. (2012). Spread of on-going changes in an immigrant language: Turkish in the Netherlands. *Review of Cognitive Linguistics*, *10*(2), 401–426.

Gelman, A., & Jennifer, H. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Gries, S. Th., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, *9*(1), 97–129.

Gries, S. Th., & Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics*, *3*. 182–200.

Gries, S. Th., & Wulff, S. (2009). Psycholinguistic and corpus linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics*, *7*, 163–186.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Jaeger, T. F., & Snider, N. E. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In *Proceedings of the 29th annual Cognitive Science Society (CogSci09)* (pp. 1061–1066), Washington, DC.

Pedersen, T. (1996). *Fishing for exactness*. Proceedings of the SCSUG 1996, Austin, TX, pp. 188–200.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633–651.

Pickering, M. J., & Holly, P. B. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, *3*(4), 136–141.

Scheepers, C., Claudine, N., Raffay, C., & Myachykov, A. (2017). The lexical boost effect is not diagnostic of lexically-specific syntactic representations. *Journal of Memory and Language*, *95*, 102–115.

Snider, N. (2009). Similarity and structural priming. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 815–820). Austin, TX: Cognitive Science Society.

Thompson, S. A., & Koide, Y. (1987). Iconicity and 'indirect objects' in English. *Journal of Pragmatics*, *11*(3), 399–406.