Psycholinguistic and corpus-linguistic evidence for L2 constructions*

Stefan Th. Gries and Stefanie Wulff
University of California, Santa Barbara / University of North Texas, Denton

In Construction Grammar, highly frequent syntactic configurations are assumed to be stored as symbolic units in the mental lexicon alongside words. Considering the example of gerund and infinitival complement constructions in English (*She tried rocking the baby* vs. *She tried to rock the baby*), this study combines corpus-linguistic and experimental evidence to investigate the question whether these patterns are also stored as constructions by German foreign language learners of English. In a corpus analysis based on 3,343 instances of the two constructions from the British component of the *International Corpus of English*, a distinctive collexeme analysis was computed to identify the verbs that distinguish best between the two constructions; these verbs were used as experimental stimuli in a sentence completion experiment and a sentence acceptability rating experiment. Two kinds of short-distance priming effects were investigated in the completion data: we checked how often subjects produced an *ing*-/*to*-/'other'-construction after having rated an *ing*- or *to*-construction (rating-to-production priming), and how often they produced an *ing*-/*to*-/'other'-construction when they had produced and *ing*- or *to*-construction in the directly preceding completion (production-to-production priming). Furthermore, we considered the proportion of *to*-completions before a completion in the questionnaire as a measure of a within-subject accumulative priming effect. We found no rating-to-production priming effects in the expected direction, but a weak effect in the opposite direction; short-distance production-to-production priming effects from *ing* to *ing* and from 'other' and *to* to *to*, and, on the whole at least, a suggestive accumulative production-to-production priming effect for both constructions. In the rating task, we found that subjects rate sentences better when the sentential structure is compatible with the main verb's collexemic distinctiveness.

Keywords: complementation, *to*, V-*ing*, constructions, patterns, second/foreign language learning, corpora, collostructions, experiments, priming, acceptability judgments

Annual Review of Cognitive Linguistics 7 (2009), **163–186**. DOI 10.1075/arcl.7.07gri
ISSN 1572–0268 / E-ISSN 1572–0276

> "There ought to be a big award for anyone who can describe exactly what makes him say 'I started to work' on one occasion and 'I started working' on another" (Quirk et al. 1974, pp. 66–7)

1. Introduction

1.1 The two complementation constructions and their problems for learners

Given the multitude of sentential complementation patterns available in English and their semantic similarity, choosing the right complementation pattern constitutes a difficulty for learners of English: even if they have acquired all the various complementation types, they may still not have mastered the lexical restrictions that come with the different complementation patterns, and consequently combine main verbs and complementation patterns that are ungrammatical or at least dispreferred in the target language (cf. Celce-Murcia and Larsen Freeman, 1999, p. 645; Schwartz and Lin Causarano, 2007). Two complementation patterns are particularly difficult to tell apart for learners of English: the verb of a main clause may either be followed by a subject-control infinitive as in (1a), or by a gerund as in (1b). We subsequently refer to the two complementation constructions as the *to*-construction and the *ing*-construction, respectively.

(1) a. People began to make strenuous efforts. (ICE-GB: W1A-012)
b. Tutors began making arrangements for sick students. (ICE-GB: W1B-018)

When being compared with each other in minimal contexts, the two complementation constructions differ more or less systematically on a number of semantic dimensions. For one, the *ing*-construction often appears to be the preferred choice to express general events, whereas the *to*-construction in the same context licenses a more specific reading, as exemplified in (2a) and (2b) (Biber et al.'s 1998, p. 758 examples).

(2) a. I tried rocking the baby gently when it cried.
b. I tried to rock the baby gently when it cried.

Secondly, the *ing*-construction denotes actuality whereas the *to*-construction denotes potentiality, as illustrated in (3) (Quirk et al.'s 1985, p. 1191 examples; cf. also Bolinger, 1968, p. 24).

(3) a. Sheila tried to bribe the jailor.
b. Sheila tried bribing the jailor.

Yet another difference concerns the temporal interpretation of the event described: the *ing*-construction invites a simultaneous interpretation in relation to the time of the utterance, whereas the *to*-construction seems to point to the future, as can be seen in (4) (Quirk et al.'s 1985, p. 1193 examples; cf. also Dixon, 1984, p. 590).

(4) a. I remembered filling out the form.
b. I remembered to fill out the form.

While these semantic tendencies may provide the learner with some rules of thumb as to the typical contexts of the two constructions, the picture is complicated by the fact that the verbs that occur in either construction are not neatly divided into two sets. As is well established for many other constructions (cf. Roland et al. 2007 for a recent data-driven approach), and as we will show below for the two complementation constructions, different verbs exhibit differently strong biases towards occurring in either construction: while some verbs nearly exclusively occur with the *to*-construction (such as *allow*, *need*, *offer*, *promise*) and others are typical of the *ing*-construction alone (such as *appreciate*, *enjoy*, *finish*, *mind*), a number of verbs may accompany both constructions more or less equally often (such as *begin*, *start*, *like*, *prefer*). Accordingly, many previous attempts to classify certain (classes of) verbs as belonging exclusively to either complementation construction need to be revisited given the evidence obtained from corpus data. To give but a few examples, in a previous analysis based on all occurrences of the two complementation patterns in the British component of the *International Corpus of English* (ICE-GB), Wulff and Gries (2004) found that contrary to previous claims (Kiparsky and Kiparsky, 1971, pp. 347f.), the *ing*-construction is not restricted to follow factive predicates (examples in case include *consider* and *avoid*); similarly, the verbs in the *ing*-construction need not be implicatives as was suggested by Givón (1990, p. 534): examples in the corpus include verbs like *recommend* and *suggest*.[1]

Yet another potential source of difficulty for learners with respect to these two complementation constructions is that the equivalents of the *to*-construction are much more common in most languages than the equivalents of the *ing*-construction (cf. Mair, 2003), and similarly, while most languages do have an equivalent to the *to*-construction, the *ing*-variant is far less common cross-linguistically (cf. Butyoi, 1977), so depending on their first language, some learners do not benefit from the possibility of positive transfer, but have to readjust their concepts of this kind of complementation in their interlanguage to a more fine-grained level.

Despite the fact that the choice between the *to*- and the *ing*-construction figures prominently in both instruction materials (cf. e.g. Werner and Nelson, 2002; McClelland and Marcotte, 2003; Frodesen and Eyring, 2007) and proficiency tests (cf. the Michigan English Language Assessment Battery), there is surprisingly little

research on this topic from the viewpoint of second language acquisition. The few contributions on this topic tend to agree that the underlying reasons for learners' difficulties with this kind of complementation resides in the grammatical representation of the two constructions: while the *to*-construction is unmarked, the *ing*-construction is marked since it does not have a complementizer position, and the gerund is only licensed by the strict subcategorization frame of the matrix verb. The fact that the *ing*-construction is so much rarer relative to the *to*-construction both within and between languages is taken to reflect this difference in markedness. Several studies indeed suggest that this difference also manifests itself in the acquisition of the two constructions such that the *ing*-construction is acquired later and less accurately (cf. Anderson, 1976 on Spanish and Persian learners, Mazurkewich, 1988 on Inuit learners, and Schwartz and Lin Causarano, 2007 on Spanish learners of English). As a matter of fact, a highly similar picture emerges for first language acquisition (cf. Pinker, 1984).

In this paper, we are concerned with advanced L2 learners' use of the two complementation constructions. In the following section, we will outline our theoretical approach and its applicability to these patterns as well as discuss some earlier work on which this study builds.

1.2 Our theoretical approach: Construction grammar

1.2.1 *Evidence for constructions in L1*

In this paper, we adopt a constructionist approach to language (cf. Goldberg, 1995, 2006). In Construction Grammar, the Saussurian concept of a symbolic unit, that is a form-meaning pair, is assumed to cover not only the level of words, but applies to constructions at all levels of semantic linguistic representation from morphemes and words to increasingly complex syntactic configurations. While many constructions are fairly non-compositional (i.e., the meanings of the parts of the construction do not add up to the meaning of that same construction, such that in *kick the bucket*, the meanings of the *kick* and *bucket* do not reveal much about the meaning of the idiomatic phrase), complex expressions also qualify as constructions if they are only sufficiently frequent and, in consequence, mentally entrenched as a holistic unit (cf. Goldberg, 2006, p. 64).

While the psychological reality of lower-level constructions such as morphemes and words is hardly ever taken into doubt, the ontological status of higher-level constructions such as sentence-level constructions requires validation. An increasing number of empirical studies, particularly on first language acquisition and syntactic priming, are more easily accounted for in a constructionist rather than a more modular framework and therefore lend credence to a construction-based account (for a recent overview, cf. Goldberg, 2006, Chapter 4).

As to language acquisition, Bates and Goodman (1997), for instance, quote several studies which have shown that morphosyntactic development and vocabulary size are highly correlated (which they show also to be true for aphasics), which supports the idea of morphemes as constructions on an equal footing as words. Another example is the work by Tomasello and colleagues, who have demonstrated in a series of studies that children's acquisition of verbal semantics beyond the holophrase stage is most adequately described in constructionist terms (cf. Braine and Brooks, 1995, Tomasello, 1998, Tomasello, 2003).

As to syntactic priming, i.e., the fact that speakers tend to reuse syntactic patterns they have encountered (comprehended or produced) before, numerous studies suggest that linguistic processing has to involve some kind of phrase structure representation (cf. Bock, 1986; Bock and Loebell, 1990; Loebell and Bock, 2003). However, Goldberg and colleagues present evidence that priming seems to involve not only syntactic, but also semantic/thematic information, which supports the conception of the psychological representation of constructions as form-meaning pairs rather than mere syntactic patterns (cf. Hare and Goldberg, 1999 and especially Chang, Bock and Goldberg, 2003).

1.2.2 *Evidence for constructions in L2*

The growing number of studies speaking in favor of the psychological reality of constructions in native speakers' language raises the question whether constructions are also part of language learners' mental lexicon, and whether (and when over the course of their interlanguage development) they fine-tune their constructional knowledge to construction-specific preferences in terms of the words that preferably occur in those constructions. Accordingly, in an earlier study (Gries and Wulff, 2005), we investigated whether argument structure constructions can also be argued to be a part of second language learners' mental lexicon, and to what extent language learners are aware of the construction-specific verb preferences of these argument structure constructions as they are observable in native speaker corpus data (cf. Gries and Stefanowitsch, 2004). We addressed the former question by conducting two experiments, a syntactic priming and a semantic sorting study; the latter question we explored by comparing the experimental data obtained with native speakers and learner data, respectively.

The first experiment investigated if and to what extent learners can be primed for argument structure constructions in a sentence completion task (the experiment was a replication of Pickering and Branigan, 1998). German advanced learners of English were asked to complete sentence fragments; some of these sentence fragments served as primes for the ditransitive construction (such as *The racing driver showed the helpful mechanic …*), others served as primes for the prepositional dative construction (such as *The racing driver showed the torn overall …*).

These primes were immediately followed by a sentence fragment that was not biased towards a completion with either construction type (such as *The angry student gave …*), so the dependent variable here was the subjects' choice of construction.

The second experiment investigated to what extent language learners' potential semantic knowledge of argument structure constructions is reflected in their grouping of sentences in a sorting task (as previously conducted with native speakers by Bencini and Goldberg, 2000). A different group of German advanced learners of English was given 16 cards with one sentence each on them and was asked to sort these cards into four piles of four cards. The sentences crossed four verbs (*cut*, *get*, *take*, and *throw*) and four argument structure constructions (the transitive, caused-motion, resultative, and the ditransitive construction). The dependent variable in this study was subjects' preferred sorting style: they could either adopt a (perceptually simpler) verb-based sorting style or a (perceptually more complex) construction-based sorting style that Bencini and Goldberg found for native speakers.

In a third step, the experimental data were compared with corpus data from (i) the ICE-GB as an L1 corpus and (ii) verb-subcategorization preferences attested in a parsed L1 German corpus (cf. Schulte im Walde, 2006).

The results of these three case studies showed that, first, learners do exhibit both the syntactic priming and the semantic sorting preferences that strongly support the assumption that constructions are part of their interlanguage lexicon. Second, the priming effects closely resemble those of native speakers of English in that they are very highly correlated with native speakers' verbal subcategorization preferences and at the same time completely uncorrelated with the subcategorization preferences of the German translation equivalents of these verbs, which rules out a mere translation-based explanation. In sum, the results indicate that German foreign language learners of English exhibit behavior that is fully in line with a constructionist account, and the results particularly emphasize the similarity of L1 and L2 with regard to the mental representation of constructions.

> "[F]urther priming studies are also needed to examine the influence of L1, L2 construction type and level of L2 proficiency on the priming effect of prior exposure to L2 constructions on written production."
> (Robinson and Ellis, 2008, p. 507)

2. Methods

As mentioned above, Gries and Wulff (2005) showed that advanced learners exhibit not only syntactic production-to-production priming effects whose verb-specific nature is congruent with that of native speaker data, but also sorting preferences

that are highly compatible with native speakers' constructional sorting patterns. We here extend the experimental design of Gries and Wulff (2005) in two ways. First, we study the nature and source of the potential priming effects in more detail by also investigating whether advanced learners of English exhibit similar priming effects

- for constructions other than the argument structure constructions tested in Gries and Wulff (2005), i.e., the two complementation constructions. This is relevant not only for the trivial reason that more evidence is better than less but also because a sceptic might argue that (part of) the priming effects we obtained then are neither syntactic nor constructional in nature but exclusively due to the surface order of semantic/thematic roles. More specifically, the sceptic would argue that subjects are in fact exclusively reacting to the ditransitives' and prepositional datives' orders of NP_{Agent}-$NP_{Recipient}$-$NP_{Patient}$ and NP_{Agent}-$NP_{Patient}$-*to*-$NP_{Recipient}$ respectively. The present study, therefore, tests for priming effects with two constructions which probabilistically exhibit semantic/functional differences but where these differences are not associated with semantic/thematic-role ordering that could result in priming effects already on its own;
- when they produce the prime themselves as in Gries and Wulff (2005), but further away from the target, i.e not immediately before the relevant target;
- when they do not produce the prime themselves but when they read it and perform a metalinguistic task on it — an acceptability judgment — immediately before the relevant target.

Second, we study the nature and strength of verb-specific constructional preferences in more detail by investigating whether advanced learners of English exhibit constructional linguistic knowledge similar to that of native speakers such that

- they complete sentence fragments not just in accordance with the construction of the prime — as in Gries and Wulff (2005) — but also (or instead of that) according to the constructional preference of the main verb in the target sentence fragment, or maybe even according to the constructional preference of the main verb in the prime sentence;
- they rate sentences' overall acceptability in a way that correlates positively with the constructional preferences of the main verbs.

The preparation of this experiment involved two steps. First, we did a corpus study to obtain the verb-specific constructional preferences for the two complementation constructions from L1 data; we explain this step in Section *2.1*. Second, we used the results of the corpus study to design an experiment, whose setup we cover in Section *2.2*.

2.1 The corpus work

We used the ICE-CUP concordance program to retrieve from the British component of the *International Corpus of English* (ICE-GB) all instances of

- one verb immediately followed by another verb;
- an auxiliary followed by a verb;
- one verb followed by *to* followed by another verb;
- one auxiliary followed by *to* followed by another verb.[2]

The resulting concordances, which comprised 45,933 hits, were then checked manually for true hits of the two complementation constructions. The main criterion for identifying true positives was the semantic constraint that the two verbs have to denote a unitary event with the first verb specifying the action denoted by the second (cf. Langacker, 1991, p. 445). Accordingly, among many other things, instances of the *going to*-future, subordinating purpose clauses (i.e., instances licensing an 'in order to'-reading), nominalizations, and auxiliary-verb sequences (auxiliary here being defined in its traditional sense) were filtered out. After removing all false hits, 480 tokens of the *ing*-construction (48 different verb types) and 2,863 tokens of the *to*-construction (98 different verb types) remained in the data sample, which comprises 120 verb types overall.

In order to determine which verbs are particularly associated with the two complementation constructions in the target language more systematically than an inspection of the raw frequencies would allow for, we computed a distinctive collexeme analysis. Distinctive collexeme analysis (DCA) is one member in the family of collostructional analyses developed by Gries and Stefanowitsch (cf. Gries and Stefanowitsch, 2004).[3] The most basic application of that family of methods is collexeme analysis, an extension of the concept of significant collocates to co-occurrences not just of two words, but of words and other linguistic elements, most notably syntactic patterns or constructions. Lexemes that are significantly associated with a construction are referred to as collexemes of that construction, where the association is quantified by means of the log to the base of 10 of the *p*-value of the Fisher Yates exact test (cf. Stefanowitsch and Gries, 2003, pp. 217–8 for justification). A DCA is an extension of collexeme analysis that specifically compares several, typically closely related or even largely synonymous constructions such as the two constructions that make up the dative alternation (in (5) and (6)) or particle placement (in (7)) etc.

(5) a. John gave Mary the book.
    b. John gave the book to Mary.

(6) a. John made Mary a sandwich.
b. John made a sandwich for Mary.

(7) a. John picked up the book.
b. John picked the book up.

The DCA identifies those lexemes in a specified slot of the construction that distinguish best between the two constructions in question, that is, those lexemes that highlight their functional differences. The extreme case would be one where a lexeme does not occur in one construction at all. These lexemes — i.e., lexemes that are significantly distinctive for either construction — are referred to as distinctive collexemes. In the present study, we compared the *to*- and the *ing*-construction this way. In accordance with previous complementation studies discussed in Section 1, the most interesting variable slot is that of the first verb. In order to test whether any given verb lemma qualifies as a distinctive collexeme of either complementation construction, the four frequencies listed in (8) need to be determined.

(8) – the token frequency of that lemma in the *to*-construction;
– the token frequency of that lemma in the *ing*-construction;
– the frequency of the *to*-construction;
– the frequency of the *ing*-construction.

For each verb lemma, these frequencies are entered into a 2-by-2 matrix to compute the *p*-value of the Fisher-Yates exact test, which is then, for ease of exposition, log-transformed to the base of ten. Accordingly, any such value that is equal to or higher than approximately 1.3 corresponds to a probability of error of exactly or less than 5%, that is, it is statistically significant; the higher the log-transformed value, the higher the verb's distinctiveness. We retrieved all relevant frequencies for all verb lemmas that were attested in either of the two constructions and computed the DCA for the above-mentioned 120 verb lemmas with Coll.analysis 3 (Gries, 2004).

(9) provides, in decreasing order of strength of attraction, the distinctive collexemes of the *ing*-construction and the *to*-construction respectively; the distinctive collexeme strengths are given in parentheses.[4]

(9) a. *keep* (76.45), ***start*** (35.23), ***stop*** (29.45), ***avoid*** (11.87), *end up* (11.87), ***enjoy*** (11.87), *mind* (11.87), *remember* (10.14), *go* (7.99), ***consider*** (5.45), *envisage* (3.38), ***finish*** (3.38), *work* (3.38), *carry* (2.53), *fancy* (2.53), *imagine* (2.53)
b. *want* (55.67), ***try*** (22.44), *wish* (5.39), ***manage*** (4.77), *seek* (4.35), *tend* (4.06), *intend* (3.67), ***attempt*** (3.19), *hope* (3.19), ***fail*** (3.09), ***like*** (3.03), *refuse* (2.98), *learn* (2.1), *plan* (1.89), ***continue*** (1.53), *afford* (1.49)

Of these distinctive collexemes, the 12 bold-faced verbs — half of them distinctive for the *to*-construction, half of them distinctive for the *ing*-construction, and within both groups half were only attested in the construction for which they are distinctive — was used as experimental stimuli for the experimental design to be described in more detail in the following.

2.2 Experimental design

After obtaining the verb-specific constructional preferences from the corpus data, we prepared a questionnaire experiment. In order to be able to address all our questions, we used a questionnaire with two different kinds of experimental tasks, a sentence completion task and an acceptability rating task. Accordingly, each subject was presented, in alternating order, with a complete (prime) sentence to rate with regard to its overall acceptability, followed by a (target) sentence fragment to be completed.

As to the rating task, the experimental items were prepared to represent the following independent variables:

- an independent categorical variable CX_PRIME: the structure of the sentences that were to be judged and that simultaneously served as primes for the completion task: either *to*-construction or *ing*-construction;
- an independent categorical variable V_PRIME_DIST: the construction for which the main verb in these prime sentences is distinctive: either the *to*-construction or the *ing*-construction (as determined by the DCA score).

The dependent variable is the acceptability judgment RATING. The subjects were asked to indicate on a scale from −3 to +3 how well-formed the (prime) sentence sounded to them, such that a totally ungrammatical sentence would be assigned a −3, an intermediate sentence a 0, and a perfectly well-formed sentence a +3. Subjects were allowed to use any number within that range, including decimals. If they could not decide on a rating for some reason, they were requested to indicate this by putting down a question mark instead.

As to the completion task, apart from the above CX_PRIME and V_PRIME_DIST, we also included the following variables:

- an additional categorical independent variable V_TARGET_DIST: the construction for which the verb provided as part of the target sentence fragment is distinctive (as determined by the DCA score): either the *to*-construction or the *ing*-construction;
- a categorical moderator variable SELF_PRIME: the construction the subjects themselves had provided to the last experimental item, which could be

  - *ing*-completion such as *going to a museum*, *offending the professor*, or *talking to his mother*;
  - *to*-completion such as *to cross the road*, *to ask the customer for his ID*, or *to laugh*;
  - 'other' such as *the race*, *coffee and cigarettes*, *immediately*, or *abruptly when they saw the end of the road*.
- a numerical moderator variable SELF_TO_RATIO: the proportion of *to*-completions out of all *to* or *ing*-completions a subject had produced up to the current sentence fragment. For example, on the first sentence completion, this value is by definition zero, since no previous completion exists. If the subject completes the first sentence fragment with a *to*-construction, then the value of SELF_TO_RATIO for the second sentence fragment is 1/1 = 1. If the subject completes the second sentence fragment with an *ing*-construction, then the value of SELF_TO_RATIO for the third sentence fragment is ½ = 0.5, etc.
- an independent numerical variable V_FREQ: the logged lemma frequency (in the British National Corpus) of the verbs in the target fragment.

The dependent variable for the completion task is RESPONSE_CODE. The subjects were not informed about the phenomenon investigated in the experiment until after its completion. Accordingly, the instructions were left as vague as possible, only informing the participants that the purpose of the experiment was to "find out which kinds of English sentences advanced learners of English produce, and to see which sentences learners consider more or less acceptable." We then asked them to complete each of the (target) sentence fragments as quickly and spontaneously as possible in such a way that the result is a grammatically correct sentence of English. Of course, this variable was coded in the same way as SELF_PRIME since one target's RESPONSE_CODE is the next target's SELF_PRIME. Note also that the alternation of rating and completion tasks means that, at any one target fragment to be completed, the subject is influenced by the his own completion from two sentences ago (i.e., the response that is recorded as RESPONSE_CODE and then becomes SELF_PRIME) and his own rating from one sentence ago (i.e., the reaction to a constructional pattern that instantiates CX_PRIME and that is recorded as a response as RATING).

Subjects received a booklet with 36 sentences we created for this experiment; we did not use sentences from the corpus data to be able to better control various sources of variability. The 36 sentences were 18 sentence completion items and 18 acceptability ratings items which were alternating in order. Out of each 18 sentences, 6 were actual experimental items, the remaining items were distractors, so each participant worked on a total of 6 completion and 6 rating items. The order of the items was pseudo-random, the only restrictions being that rating

sentences which served also as potential primes for a sentence completion had to directly precede that sentence fragment, and that experimental pairs were always interrupted by at least one distractor pair. At the bottom of the instructions, it was emphasized that it is crucial that subjects complete all sentences only in the given order, and not to skip any sentence.

94 subjects (70 female and 24 male) participated in the experiment. All participants were native speakers of German who studied English linguistics in the lower division at the Friedrich-Schiller University of Jena and were, therefore, familiar with the basic logic of acceptability judgments, though of course not with linguistic accounts of the two complementation constructions. Their average age was 21.7 years (standard deviation: 2.36 years), and their English language education averaged around 11 years (standard deviation: 1.89 years), so they formed a highly homogeneous group of advanced learners of English.

3. Results

In this section, we discuss the results of both parts of the experiment; Section *3.1* is concerned with the acceptability ratings while Section *3.2* covers the completion task.

3.1 Experiment part 1: Sentence acceptability rating

We obtained 556 ratings from 94 subjects; for eight stimuli, subjects provided no responses. In order to determine which variables influenced the rating of the primes most, we first computed a linear model, which included RATING as the interval-scaled dependent variable and CX_PRIME and V_PRIME_DIST as well as their two-way interaction as independent categorical variables. This model yielded a significant overall fit ($F = 15.15$; $df = 3, 552$; $p < 0.001$), but it turned that the model has only a modest explanatory power (adj. $R^2 = 0.07$) and that only the interaction between CX_PRIME and V_PRIME_DIST is significant. This interaction, which accounts for just about all the explanatory power, is represented both in Table 1 and, from both complementary perspectives, in the interaction plots in Figure 1, which shows the mean ratings plus bars corresponding to their standard errors; the means whose standard errors overlap are not significantly different according to Tukey's HSD tests.

It is obvious that the ratings in the interaction correlate strongly with what one would expect from the DCA: when the prime was an *ing*-construction and the verb in that prime was a verb which the DCA characterized as distinctive for the *ing*-construction, then the ratings were much higher than when either the prime

Table 1. Means: CX_PRIME × V_PRIME_DIST → RATING

		V_PRIME_DIST		
		ing	*to*	pooled
CX_PRIME	*ing*	0.989	–0.289	0.35
	to	0.003	0.801	0.397
	pooled	0.479	0.268	0.374

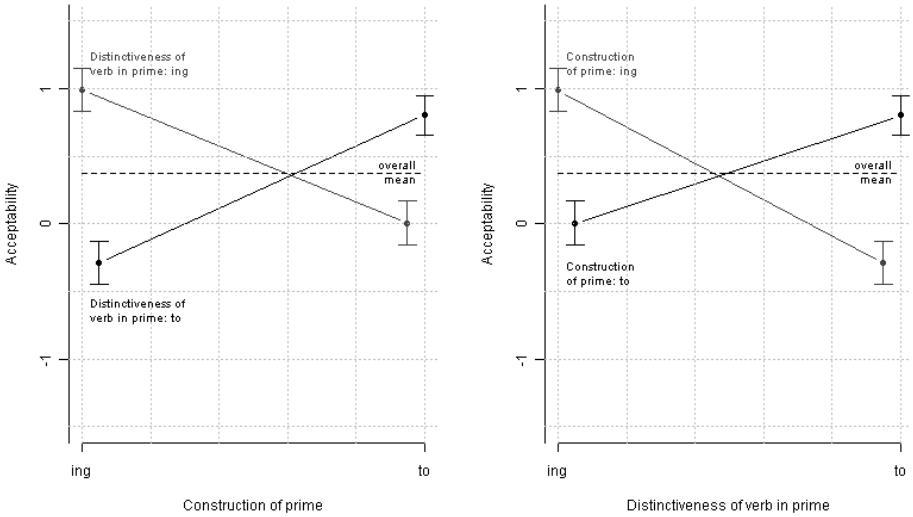

Figure 1. Interaction plots: CX_PRIME × V_PRIME_DIST → RATING

was a *to*-construction or the verb was distinctive for a *to*-construction; the same is true of the *to*-construction.

3.2 Experiment part 2: Sentence completion

We obtained 560 sentence completions from 94 subjects. Of the sentence completions, we discarded the 'other'-completions, which left us with 193 *ing*-responses and 176 *to*-responses (by altogether 92 subjects). In order to determine which variables influenced the choice of construction in the sentence completion most, we first computed a maximal logistic regression model. This model included RESPONSE_CODE as the binary dependent variable and CX_PRIME, V_PRIME_DIST, V_TARGET_DIST, SELF_PRIME, SELF_TO_RATIO, and V_FREQ as well as all their two-way and three-way interactions as predictors.

This maximal model yielded a good and significant overall fit, but nearly all interactions turned out to be insignificant. We therefore performed a model selection process in which we successively eliminated the non-significant predictors

with the highest *p*-values until we arrived at a final minimal adequate model. In this process, we eliminated all three-way interactions, all but one two-way interaction, two individual independent variables, and we conflated the one levels SELF_PRIME: *other* and SELF_PRIME: *to* into a new level SELF_PRIME: *not-ing* because they did not differ significantly from each other ($p = 0.12$).

This final model is highly significant (log-likelihood $\chi^2 = 128.48$; $df = 5$; $p < 0.001$) and has both a reasonable explanatory power (Nagelkerke's $R^2 = 0.39$) and a good predictive accuracy ($C = 0.82$, Somer's $D = 0.64$); the exact classification accuracy achieved by the model is 277/369≈75.1%, which is 50% better than the baseline accuracy. To see where the explanatory accuracy and predictive power come from, consider the parameters represented in Table 2, where the odds ratios provided reflect the change in odds to get a sentence completion with *to*.

Table 2. Results of the final binary logistic regression model

Factor	*Wald* χ^2	*df*	*p*	regression coeff.	odds ratio
V_TARGET_DIST: *to*	22.89	1	< 0.001	25.96	$1.88 \cdot 10^{11}$
V_FREQ	10.13	1	0.001	1.52	
SELF_PRIME: *ing*	12.42	1	< 0.001	−0.96	0.38
CX_PRIME: *to*	3.96	1	0.046	−0.5	0.61
V_TARGET_DIST: *to* × V_FREQ	19.59	1	< 0.001	−2.36	

The results clearly show that the verb's constructional preference in the sentence completion target fragment V_TARGET_DIST has by far the strongest influence on the complementation choice: when the verb in the target is distinctive for the *to*-construction as opposed to the *ing*-construction, this hugely increases the odds of another *to*-construction; the variable that is temporally closest to where the subject will choose a construction has the strongest effect.

Then, there is an effect of V_FREQ: the more frequent the verb lemma, the more likely a *to*-construction becomes. However, this variable participates in a significant interaction so we must return to it again below.

We can also observe a kind of a short-distance production-to-production priming effect: SELF_PRIME, a subject's last completion, has the expected effect on the immediately following completion. The regression coefficient for SELF_PRIME: *ing* is negative, which means that after the subjects produced an *ing*-construction, they are less likely to produce a *to-construction* on the next target than if they did not produce an *ing*-construction. This situation is best understood graphically, so cf. the left panel of Figure 2.

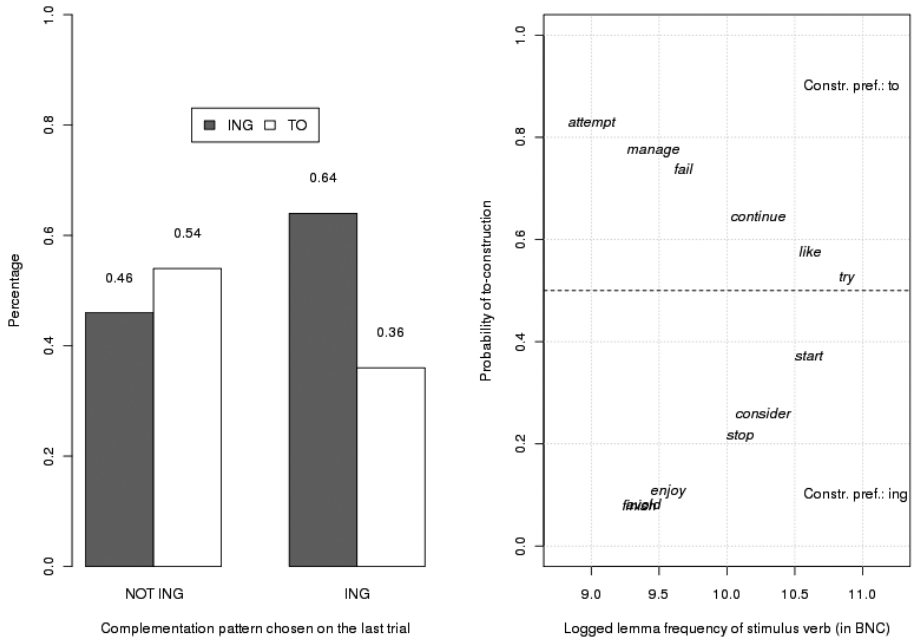

Figure 2. SELF_PRIME → RESPONSE (left panel) and V_TARGET_DIST × V_FREQ → RESPONSE (fitted probabilities from the interaction; right panel)

The last significant main effect is that of CX_PRIME: there is a weak effect such that if the construction in the prime to be rated was a *to*-construction, then another *to*-construction is a bit less likely, which is unexpected: if any effect that would be strong enough to maybe override V_TARGET_DIST was to be expected at all, the reverse direction of the effect would have been expected. Thus, the experimental sentences that were rated correlate *negatively* with the pattern the subjects use on the next fragment to be completed, which means that there is no priming from the metalinguistic judgment task to the sentence completion task.

The final significant predictor is the interaction between V_TARGET_DIST and V_FREQ. It is represented in the right panel of Figure 2. On the *x*-axis, we show the V_FREQ, on the *y*-axis the proportion of *to-constructions.* The verbs are plotted into the graph at their predicted probability of *to*-constructions.[5] The main effect of V_TARGET_DIST is reflected by the fact that there is a clear division between the two constructions that the verbs are predicted to occur in. The interaction, however, is represented by the fact that as the lemma frequencies of *to*-preferring verbs increase, fewer *to*-constructions are produced by the subjects, and at the same time, as the lemma frequencies of *ing*-preferring verbs increase, more *to*-constructions are predicted. While this appears to be a somewhat interesting result, the predictive power of the interaction alone is negligible and we are now

learning towards thinking that this basically only means that the more frequent a verb is, the more it is likely to admit both constructions more equally.

Let us now finally turn to one insignificant predictor which may be an interesting variable to control for in future studies. This effect, which in our small dataset admittedly did not reach standard levels of significance — it was the very last predictor to be eliminated in the model selection process — is suggestive evidence for what one might refer to as a within-subjects accumulative priming effect: the larger the proportion of *to*-completions that a subject provided in all their completions, the higher the likelihood of an additional *to*-completion became. Figure 3 represents this correlation graphically in the format of a spinogram for the pooled data of all subjects. On the *x*-axis, we portray the proportion of *to*-completions before a completion at item *I* in the questionnaire; onto the *y*-axis we plot the completion chosen for *I*. It is plain to see that while the proportion of *to*-completions at the beginning of a questionnaire is a little more than 40%, the more *to*-completions a subject has already produced in the experiment in percent, the larger the proportion of further *to*-completions becomes.

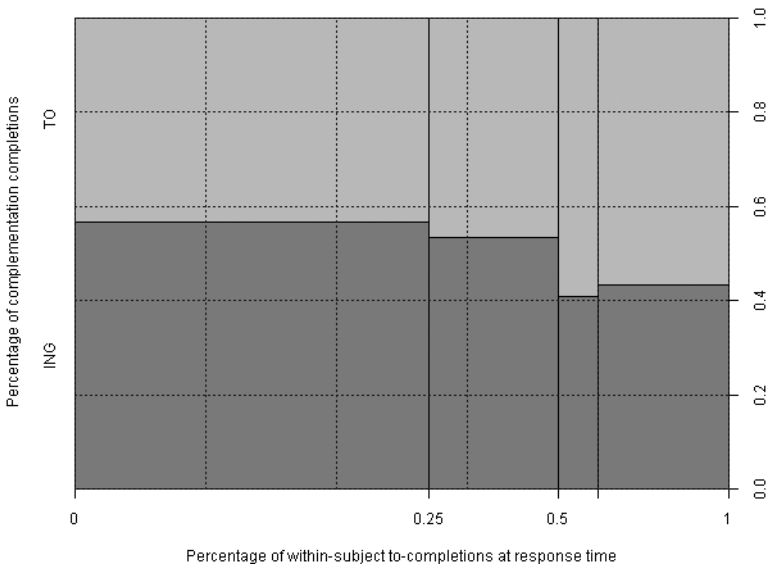

Figure 3. Spinogram of RESPONSE_CODE × SELF_TO_RATIO

The following section briefly recapitulates and then discusses the results as well as their implications and provides some conclusions.[6]

4. Discussion

4.1 Methodological conclusions

There are two kinds of methodological implications we would like to highlight. First, while the corpus data analyzed in Section *2.1* only play a subordinate role in this paper, as so often before, they paint a much more varied picture than is usually found in instruction materials and in theoretically-oriented research. Instead of neatly delineated verb classes, a large corpus analysis based on several thousand manually-identified and coded matches results in a much greater degree of overlap and 'messiness' in the data, with some verbs exhibiting strong probabilistic tendencies. In addition and more specifically, the method of DCA, which has proven useful in a variety of settings — the syntax-lexis interface, syntactic priming in L1 and L2, variety differences, differences between L2-proficiency groups, etc. — also allows us to identify groups of verbs which produce significant differences in the experiments (which in turn, is an experimental validation of the DCA itself).

Beyond that, the (non-significant) tendency for accumulative within-subject production-to-production priming effect observed in the present study could be important. While the effect was only marginal, it was not only visible overall, but more interestingly, it was differently strong for different subjects. In addition, it will also be necessary to explore to what degree different verbs exhibit different behaviors. We know of hardly any studies that have incorporated these or similar (random) effects, but the results suggest that some such mode of analysis may be usefully included in many experimental designs or even be applied in retrospect to already published data. For example, the new family of methods known as mixed-effects or multilevel modeling would be a natural starting point for follow-up analyses.[7]

4.2 The experimental results

Let us begin by recapitulating our main findings with regard to the priming effects and the verbs' constructional preferences. With regard to the former, we found

- no metalinguistic-rating-to-production priming in the expected direction, but a weak effect in the opposite direction;
- short-distance production-to-production priming: *ing* decreases the likelihood of *to;*
- weak/insignificant but suggestive accumulative within-subject long-distance production-to-production priming from *to* to *to* and *ing* to *ing*.

With regard to the latter, we found

- in the completion task: the verbs' collexemic preferences in the sentence fragment have the strongest and the expected effect on sentence completion;
- in the rating task: subjects rate sentences better when the sentential structure is compatible with the main verb's collexemic distinctiveness.

On the whole, the results for the two are in agreement with those of Gries and Wulff (2005). The two patterns studied here are not argument structure constructions and, therefore, cannot be reduced to semantic/thematic role ordering effects, and they yield results that provide strong support for their having attained some kind of constructional status for the L2 learners: both patterns exhibit collostructional preferences and priming effects of various kinds. However, as we summarized above, the picture is also more varied: (i) the metalinguistic rating task does not yield the same kind/direction and amount of priming as in earlier studies, in which subjects actually produced the priming structure themselves, and (ii) the production-to-production priming is in the expected direction — since *ing*-constructions are significantly primed by *ing*-constructions in the previous target (compared to *other* and *to*) — but the ratio of *to*-constructions after *to*-constructions does not differ significantly from that of *to*-constructions after *other*-constructions.

At present, we can only engage in some *post hoc* speculation as to the exact reasons for some of the unexpected results. On the one hand, the absence of rating-to-production priming may reflect a task effect. First, the metalinguistic rating task may result in the advanced language learners weighing different constructional alternatives so that the completion of the target is in fact influenced by the learner having processed both the *to*- and the *ing*-construction to arrive at a rating. This is indeed quite likely because, as we have pointed out above, the issue of *to*- vs. *ing*-complementation is one that figures prominently in many instructional manuals and is, thus, deeply entrenched in every German learner's mind. Since the learners in Gries and Wulff (2005) had to perform no such rating task, such effects were not expected and, indeed, not obtained.

Second, the facts that (i) *to*-constructions are equally frequent found after *to*- and after *other*-primes but (ii) *ing*-constructions exhibit priming may be explained as follows: As to (i), *to*-constructions have a close translational equivalent in German and are therefore likely to be inserted as a quasi-default when the prime does not evoke a particular complementation construction; hence subjects use *to* both after *to* and after *other* (i.e., in general). Also, since the subjects were explicitly addressed as advanced learners of English, there may have been a social desirability bias such that the learners wanted to portray themselves as commanding a wide variety of patterns; some *prima facie* evidence for this assumption is the fact that many sentence completions were longer and more complex than one would expect

if subjects had just tried to get through the questionnaire as quickly as possible. Arguably, this may have led them to actually avoid repetitions in particular with the *to*-construction, which may not have a particularly idiomatic feel to them for exactly the reason that there is a close German equivalent. As to (ii), since *ing*-constructions do not have a similarly close equivalent, they are learned later and more consciously and may, thus, more likely benefit from priming.

One may wonder why we do not interpret what superficially appears to be a lack of *to*-to-*to* priming as evidence against the constructional perspective we have adopted. The are several reasons for this. First, as we discussed above, the *to*-pattern *is* still significantly different from *ing* — it is just not significantly different from *other*. Second, although the priming effect is not as clear as it could have been, one must not forget that the *to*- and the *ing*-pattern still qualify as constructions in the sense of, say, Goldberg (2006, p. 5): the patterns are obviously formally different, and they are functionally different in the sense of (i) having predictably different meanings in at least some contrastive minimal pairs and (ii) exhibiting lexicogrammatical associations (the verb-specific preferences observed in the DCA) of exactly the same kind that have been observed for a large variety of patterns whose constructional status is beyond doubt.

Third, we know from Chang, Bock, and Goldberg (2003) that some priming effects cannot be explained with recourse to a purely syntactic, or constituent structure, notion of priming anyway to accommodate effects that are compatible with the assumption of links between thematic roles and structural positions. Thus, since the *to*- and the *ing*-pattern meet the definition of constructions *and* since we need an account of priming that is compatible with constructions anyway, it is actually more parsimonious to assume constructional priming, which, since constructions are pairings of form and meaning, can accommodate both syntactic and semantic/thematic aspects of priming.

A final reason is the accumulative within-subject production-to-production priming mentioned in the previous section. The above argumentation regarding the absence of priming of the *to*-construction largely involved relatively conscious learner strategies. Given the close temporal proximity between individual prime and target pairs, learners may well consciously note parallelisms and exhibit the above avoidance strategies. The accumulative priming effect, by contrast, taps into much less accessible knowledge and processes since the source of the priming is distributed across the complete questionnaire. The fact that at least some weak kind of priming is observed provides some support for the constructional status of the *to*-pattern.

With regard to the verb-specific constructional preferences, the results of both the rating and the completion task are very similar to each other as well as to the earlier findings in Gries and Wulff (2005). Both add to the growing body

of literature testifying to item-specific knowledge in general as well as language learners' verb-specific knowledge in particular (cf. also Gries, 2006, Section 2). It is interesting to note, however, that the verb's preference in the prime had no effect whatsoever on the sentence completion: rather, the preference of the verb in the target fragment appears to overpower all verb preferences from previous material, and even most of the constructional priming from the previous construction. This, as well as the priming results, is fully compatible with what we know from first language acquisition, where analyses of corpus data reveal similar strong lexico-constructional associations (cf. Kidd, Lieven and Tomasello, 2006); thus, the present paper is compatible with a constructional approach to language. We feel that these strong, but nevertheless probabilistic, effects are best accounted for in a probabilistic approach to acquisition and learning (such as, e.g., exemplar theory; cf. Ellis, 2008) and hope that our study has taken a small step to arrive at a deeper understanding of learners' knowledge and processing.

Acknowledgments

We thank Beate Hampe and Holger Diessel for their invaluable assistance in collecting the experimental data for us, Nick C. Ellis for helpful discussion, and two anonymous reviewers as well as the audiences at L.A.U.D. 2008, AILA 2008, AAAL 2009, and the Applied Linguistics colloquium at UCSB for comments at various different stages of the project. The usual disclaimers apply.

Notes

* The order of authors is arbitrary.

1. On the other hand, some previous analyses are only insufficiently precise: Huddleston and Pullum (2002, p. 1241) assume that verbs like *bother*, *intend*, *plan*, and *attempt* behave indifferent with regard to the two constructions. As is obvious from the corpus-based results of the present study (cf. Section *2.1*), quite the contrary holds: *bother* is significantly distinctive for the *ing*-construction, and *intend* and *attempt,* and *plan* are significantly distinctive for the *to*-construction.

2. For the initial retrieval, we had to include auxiliaries because in the ICE-GB tagging scheme, verbs like *begin* and *start* are also sometimes tagged as auxiliaries.

3. Unless indicated otherwise, all retrieval operations, computations, and graphs were performed with R for Windows 2.8.1 (R Development Core Team 2009).

4. Note that even verbs that occur only in one construction may result in different DCA values, which is due to their different overall frequencies in the corpus (cf. Gries and Stefanowitsch, 2004 for discussion and exemplification).

5. This prediction is only based on the interaction in question.

6. We also ran one model selection process in which we included a variable that represented whether the verb in the target fragments was not just distinctive for one of the two constructions, but also (near-)categorically restricted to that construction. The numerical results of course changed, but the conceptual implications with regard to priming, verb-specific preferences etc. just discussed were the same.

7. Initial explorations along these lines suggest that the results of our logistic regression stay very much the same: the predictors that survive the model selection process are the same, their signs and strengths are comparable, and no by-subject adjustment is necessary. However, by-verb adjustments play a significant role in the prediction of the completion data and boost the classification accuracy from the above 75.1% to 303/369 ≈ 82.1% ($C = 0.9$, Somer's $D = 0.79$). Figure (i) below represents the sizes of these adjustments on the *y*-axis (jittered along the *x*-axis to avoid overplotting) and shows that the model needs to make the largest adjustment for *try* and *avoid* as well as *continue* and *like*. Findings like these need to be studied in more detail …

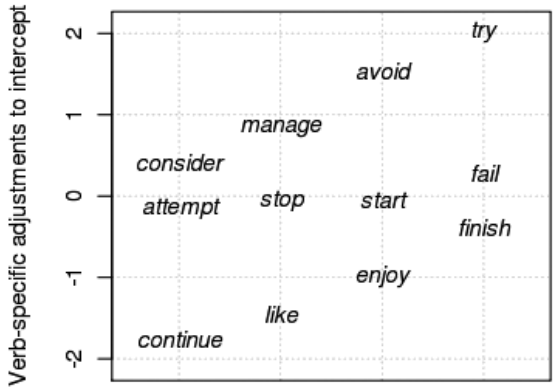

Figure (i). Verb-specific adjustments to the intercept in a generalized linear mixed-effects model

References

Anderson, J.I. (1976). A comparison of the order of difficulty of English sentential complements between native speakers of Spanish and native speakers of Persian. Paper presented at the Los Angeles Second Language Research Forum, UCLA.

Bates, E. & J. Goodman. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, 12(5/6), 507–86.

Bencini, G.M.L. & A.E. Goldberg. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language,* 43(4), 640–51.

Biber, D, Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1998). *Longman Grammar of Spoken and Written English*. London: Longman.

Bock, J.K. (1986). Syntactic persistence in language production. *Cognitive Psychology,* 18(1), 355–87.
Bock, J.K. & H. Loebell. (1990). Framing sentences. *Cognition,* 5(1), 1–39.
Bolinger, D.L. (1968). *Aspects of Language*. New York: Hartcourt, Brace, and World.
Braine, M.D.S. & P.J. Brooks. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello & W.E. Merriman (Eds.), *Beyond Names for Things: young children's acquisition of verbs* (pp. 353–76). Hillsdale, NJ: Lawrence Erlbaum.
Butoyi, C.A. (1977). The accuracy order of sentential complements by ESL learners. Unpublished M.A. thesis, UCLA.
Celce-Murcia, M. & D. Larsen-Freeman. (1999). *The Grammar Book: An ESL/EFL teacher's course*. Boston, MA: Heinle and Heinle.
Chang, F., Bock, J.K., & Goldberg, A.E. (2003). Can thematic roles leave traces of their places? *Cognition,* 90(1), 29–49.
Dixon, R.M.W. (1984). The semantic basis of syntactic properties. *BLS* 10, 583–95.
Duffley, P.J. (1995). Defining the potential meaning of the *-ing* form in English. *Langues et linguistique,* 21: 1–11.
Ellis, N.C. (2008). The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N.C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 372–405). London, New York: Routledge.
Frodesen, J. & J. Eyring. (2007). *Grammar Dimensions*. 4th ed. Boston, MA: Thomson Heinle.
Goldberg, A.E. (1995). *Constructions: a construction-grammar approach to argument structure*. Chicago/London: University of Chicago Press.
Goldberg, A.E. (2006). *Constructions at Work: the nature of generalization in language*. Oxford: Oxford University Press.
Gries, S.Th. (2004). *Coll.analysis 3. A program for R for Windows*.
Gries, S.Th. (2006). Some proposals towards more rigorous corpus linguistics. *Zeitschrift für Anglistik und Amerikanistik,* 54(2), 191–202.
Gries, S.Th. & A. Stefanowitsch. (2004). Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics,* 9(1), 97–129.
Gries, S.Th. & S. Wulff. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. *Annual Review of Cognitive Linguistics,* 3, 182–200.
Hare, M.L. & A.E. Goldberg. (1999). Structural priming: Purely syntactic? In M. Hahn & S.C. Stones (Eds.), *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (pp. 208–11).
Huddleston, R.D. & G.K. Pullum. (2005). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
Kidd, E., Lieven, E. & Tomasello, M. (2006). Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development,* 21(2), 93–107.
Kiparsky, P. & C. Kiparsky. (1971). Fact. In D. Steinberg & L.A. Jakobovits (Eds.), *Semantics: an interdisciplinary reader in philosophy, linguistics, and psychology (pp. 345–69)*. Cambridge: Cambridge University Press.
Langacker, R.W. (1991). *Foundations of Cognitive Grammar*. Vol. II. Stanford, CA: Stanford University Press.
Loebell, H. & J.K. Bock. (2003). Structural priming across languages. *Linguistics,* 41(5), 791–824.
Mair, C. (2003). Gerundial complements after *begin* and *start*: Grammatical and sociolinguistic factors, and how they work against each other. In G. Rohdenburg & B. Mohndorf (Eds.),

Determinants of Grammatical Variation in English (pp 347–77). Berlin, New York: Mouton de Gruyter.
Mazurkewich, I. (1988). The acquisition of infinitive and gerund complements by second language learners. In S. Flynn & W. O'Neil (Eds), *Linguistic Theory and Second Language Acquisition* (pp. 127–43). Dordrecht: Kluwer.
McClelland, L.D. & P.H. Marcotte. (2003). *Writing Matters! Introduction to writing and grammar*. New York: McGraw Hill.
Pickering, M.J. & H.P. Branigan. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language,* 39(4), 633–51.
Pinker, S.A. (1984). *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.
Quirk, R. (1974). *The Linguist and the English Language*. London: Arnold.
Quirk, R., Greenbaum, S., Leech, G., & Svartvik, S. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
R Development Core Team. (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
Robinson, P. & Ellis, N.C. (2008). Conclusion: cognitive linguistics, second language acquisition and L2 instruction — issues for research. In P. Robinson & N.C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition (pp. 489–545)*.
Roland, D., Dick, F. &, Elman, J.L. (2007). Frequency of basic English grammatical structures: a corpus analysis. *Journal of Memory and Language,* 57(3), 348–79.
Schulte im Walde, S. (2006). Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics,* 32(2), 159–94.
Schwartz, M. & P.L. Causarano. (2007). The role of frequency in SLA: an analysis of gerunds and infinitives in ESL written discourse. *The Arizona Working Papers in SLA and Teaching* 14, 43–57.
Stefanowitsch, A. & S.Th. Gries. (2003). Collostructions: On the interaction between verbs and constructions. *International Journal of Corpus Linguistics,* 8(2), 209–43.
Tomasello, M. (1998). The return of constructions. *Journal of Child Language,* 25(2), 431–42.
Tomasello, M. (2003). *Constructing a Language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Werner, P.K. and J.P. Nelson. (2002). *Mosaic 2 Grammar*. 4th ed. New York: McGraw-Hill.
Wulff, S. & S.Th. Gries. (2004). *Prefer to construe* vs. *prefer construing*: a corpus-linguistic perspective on non-finite sentential complementation (Paper presented at Current Trends in Cognitive Linguistics, University of Hamburg, 11 December 2004).

Author's address

Dr. Stefan Th. Gries
Department of Linguistics
University of California, Santa Barbara
Santa Barbara, CA 93106-3100
USA

stgries@linguistics.ucsb.edu

About the authors

Stefan Th. Gries is currently Associate Professor in the Department of Linguistics at UC Santa Barbara. He has been working in the fields of cognitive linguistics and quantitative corpus linguistics and is particularly interested in morphological and lexico-grammatical phenomena as well as statistical methods in linguistics.

Stefanie Wulff is currently Assistant Professor in the Department of Linguistics and Technical Communication at the University of North Texas. She has been working in the fields of second and foreign language acquisition, phraseology, and corpus linguistics.