CHAPTER 3

Corpus-driven methods for assessing accuracy in learner production

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

Adopting the perspective of Ellis's (2007) Associative-Cognitive CREED, this chapter proposes a measure of accuracy in learner production that is based on conditional probabilities. More specifically, we develop a definition of accuracy that involves 'the proficient selection of constructions in their preferred constructional context in a particular target genre'. Comparing this approach to previous work on linguistic units larger than the word, we discuss how this definition (i) does away with a strict separation of lexis and grammar, shifting the focus to interactions between constructions; (ii) embraces various aspects of accuracy (phonology, morphology, lexis, etc.) instead of being restricted to target-like vocabulary choice alone; and (iii) reflects our understanding of native-like proficiency as a gradual, probabilistic phenomenon that transcends a native-nonnative speaker divide. We then exemplify this measure in two small case studies using lexico-grammatical association patterns from L1 and L2 corpora and discuss implications of the theoretical perspective and the empirical measure for task design.

Introduction

Accuracy is usually very widely defined as the native-like use of different linguistic features, including pronunciation, grammatical morphemes, and maybe most prominently, adequate vocabulary choice. Commonly labeled as a primarily grammatical phenomenon, it is often contrasted with fluency as its pragmatic counterpart. A typical example is Byrd's (2005) definition:

> In most uses, accuracy refers to "grammatical accuracy" but other areas of language use can be involved too: spelling and/or pronunciation. Fluency implies the ability to easily understand and participate in communication, generally spoken, in the person's second language. (p. 551)

Byrd goes on to note that rather than being in direct opposition, however, recent research suggests an intricate interplay between the two. An even more complicated picture presents itself in various recent SLA studies referred to as Complexity-Accuracy-Fluency (CAF) studies, which define general language proficiency as the complex interplay of all three dimensions (see Wolfe-Quintero et al., 1998; Ellis & Yan, 2004; Larsen-Freeman, 2006; and Housen & Kuiken, 2009 for an excellent summary of ongoing issues regarding the definition of CAF).

We take this line of reasoning one step further and propose a definition of accuracy that accommodates recent findings concerning the interplay between accuracy and fluency, and which, moreover, is compatible with contemporary linguistic theorizing inside and outside SLA. In recent studies in theoretical linguistics, psycholinguistics, and corpus linguistics, the long-held dichotomy of grammar and lexis has come under serious attack. One such framework that basically discards this distinction altogether is Construction Grammar, and we describe some relevant assumptions here in Section 2.

We then devote Section 3 to a brief summary of three strands of research in SLA that are, if not explicitly constructionist in nature, highly compatible with such an approach. With these findings in mind, we propose our definition of accuracy in Section 4, and discuss a corpus-linguistic method that can be used as a measure of our definition. In Section 5, we present two case studies to illustrate the potential of this approach to accuracy. In Section 6, we discuss some implications for issues of task design, particularly with regard to task complexity, before we round off the chapter with some general conclusions.

A constructionist perspective on language

In this paper, we adopt a constructionist approach to language (cf. Goldberg, 1995; 2006). In Construction Grammar, constructions are defined as form-meaning pairs that exist at all levels of linguistic representation:

> Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (Goldberg, 2006, p. 5)

In this sense, the notion of construction embraces, in addition to words and morphemes, all kinds of more or less formally fixed, schematic (i.e., lexically filled or unfilled), and semantically transparent expressions. These have formerly been given various names in the SLA literature and elsewhere, including *prefabricated patterns*, *routines*, *chunks*, *free combinations*, (*restricted) collocation*s, *idioms*, and so on – in Construction Grammar, we can describe all of these expressions in one common framework.

The branch of Construction Grammar we follow here is a non-generative theory in which any complex utterance is a combination of various constructions. Goldberg (2006, p. 10) provides the example of the sentence *what did Liza buy Zach?*, which involves (at least) the following constructions: the words *Liza*, *buy*, *Zach*, and *what*; a ditransitive construction; a question construction; a subject-auxiliary inversion construction; a VP construction; and an NP construction. Constructions are freely combinable as long as their specifications are compatible with each other. In cases of direct conflict, the resulting sentence will either be judged ill-formed (think, for example, of a learner combining the subject-auxiliary inversion construction with a non-question construction) or else lower-level specifications will override higher-level specifications.

With regard to ill-formedness, it furthermore needs to be emphasized that Construction Grammar is a usage-based approach: what is considered well-formed (or, in other words, accurate) is often a matter of degree, and more often than not a function of (conditional) probability/frequency of usage. Crucially, the well-formedness of a complex utterance is correlated to some degree with the absolute frequency of every construction that makes up the utterance (such that generally speaking, using frequent words and other constructions will most likely result in an acceptable utterance), but even more so with the frequency with which the constructions in question are used *together*. In other words, a major correlate of well-formedness are the *conditional probabilities* of pairs (or even larger clusters) of constructions. To give a simple example, *give* is a highly frequent verb in English, which can occur in both the ditransitive (*Steffi gave the squirrel some bread*) or the prepositional dative construction (*Steffi gave some bread to the squirrel*). While both combinations are grammatical, native speakers (NS) use the former combination considerably more often than the latter. Consequently, the *conditional frequency/probability* of the ditransitive is much higher than that of the prepositional dative when the verb is *give*.

A Construction Grammar approach has the following implications for language acquisition: there is no fundamental distinction between words and the grammatical rules to combine them properly. Instead, accurate mastery of a language entails the acquisition of constructions at different levels of complexity and schematization, as well as knowledge of the probabilistic tendencies underlying their target-like combination. Research in first language acquisition (Tomasello, 2003) has gathered substantial support in favor of this view; in the following, we turn to supporting studies in second language acquisition.

Previous research

L2 production research beyond the word

Early research on L2 production was far from a constructionist perspective, mainly because various concepts were not sufficiently differentiated: *what* is being acquired

(words vs. larger routines), *how* the linguistic input is being processed (analytically vs. holistically), and *in which form* it is stored (analytically vs. holistically); not to speak of the potential impact of the learning *environment* (naturalistic vs. classroom-based) and *instructional style* (explicit vs. implicit) (Weinert (1995) provides an excellent review of these parameters). In fact, the field is only beginning to disentangle these concepts and assess their individual contributions to L2 proficiency (Ellis, 1994; Norris & Ortega, 2000). Most important in the present context is the first dichotomy, words vs. larger routines or patterns, which reflects a view of language in which lexis and syntax/grammar are two separate components of the (inter)language system.

Accordingly, early research into L2 production beyond the word mostly looked for what Brown (1973) referred to as *prefabricated routines*, that is, unanalyzed multiword expressions with a particular pragmatic function. Maybe also due to Brown's influence in the field, most studies focused on children acquiring a second language. Of central concern was the question if, and to what extent, evidence for such prefabricated routines would reflect a *gestalt mode* or *expressive* learning strategy, where children start out with these prefabricated routines before breaking them down into their component parts, as opposed to using an *analytical* or *referential* learning strategy, whereby children combine words into increasingly larger units. The results of these early studies were inconclusive (cf. Krashen & Scarcella (1978) for discussion). Hatch (1972), for instance, examined production data from a 4-year old Chinese boy learning English and found evidence for both learning strategies running in parallel. Hakuta (1974) drew a sharp distinction between such prefabricated routines and what he called *prefabricated patterns*, which were defined not as wholly fixed phrases, but segments of sentences which operate in conjunction with a movable component. While Hakuta (1976) presented some evidence from a 5-year old Japanese learner of English for learning through rote memorization of such patterns, Wagner-Gough (1975) investigated the L2 production of a young boy, Homer, and concluded that prefabricated patterns did, however, apparently not transfer into creative language use, suggesting a minor role of prefabricated language in the acquisition process. Maybe the most comprehensive analysis at the time was Wong-Fillmore's (1976) dissertation, in which she tracked the L2 acquisition of five kindergartners. She argued that children start out with prefabricated patterns and only later in the acquisition process decompose these patterns into their constituent parts for rule formation and, ultimately, creative use.

Early research on adult L2 acquisition was even more scarce (for a comprehensive overview, see Wray (2002, p. 172–198)). Researchers concurred that while it is true that adult learners seem to acquire prefabricated routines to some extent, unlike children, this knowledge does not further grammatical development. One example is a study by Hanania and Gradman (1977) of Fatmah, a NS of Arabic learning English, who was 19 years old at the time and had had only very little schooling in her L1. Fatmah used routines tied to specific pragmatic situations, but ad hoc attempts to have her decompose these routines into their constituent patterns were largely unsuccessful. Shapira (1978) and Schumann (1978), working with L2 learners from different L1

backgrounds, also found only little evidence for prefabricated language or a facilitating effect of knowledge of prefabs for acquisition in general. Schmidt (1983) found that his learner Wes used prefabricated routines much more than any of the other learners, but also conceded that while Wes's extensive knowledge of routines gave him some fluency, it did not improve his grammar competence. Looking into the role of prefabricated language in classroom instruction, Ellis (1984) found that his subjects learned and used various types of memorized formulas and scripts, some of which were later used for syntactic development. However, he pointed out that there was considerable learner variation. In a cross-sectional study of the acquisition of routines in the L2 classroom, Scarcella (1979) concluded rather pessimistically that generally, adults have "difficulty acquiring very common routines" (p. 84). Accordingly, Krashen and Scarcella (1978, p. 298) recommended not encouraging adult L2 learners to focus on prefabricated language because "[t]he outside world for adults is nowhere near as predictable as the linguistic environment around Fillmore's children was".

The first to call the categorical distinction between vocabulary and syntax into question from an acquisition/learning perspective, although they may not have been aware of that at the time, were Pawley and Syder (1983). They pointed out that there is a fundamental qualitative difference between *native-like fluency*, the ability to speak fluently in a second language, and *native-like selection* (or idiomaticity), the ability to select the right words in their proper contexts. In fact, Palmer (1933, p. 8), examining second language learners' use of verb-object combinations, had already drawn attention to the problem of native-like selection 50 years earlier when he noted how learners depend on explicit instruction on the matter:

> ...Without such information the learner tends to form such combinations by guess work or the analogy of his mother tongue, and we can imagine him coining such unusual expressions as
> To make a question
> To perform a favour
> To do trouble
> To keep patience ...

This distinction between fluency and native-like selection explains the apparent contradictions in Wes's language production. It also suggests that the proper use of prefabricated language is most likely to be expected only at an advanced level of general language proficiency: a learner first *needs* to acquire simple constructions alongside the complex constructions serving as syntactic frames before they can begin to explore which words *prefer* to go into which frames.

Several studies in the 1980s supported this position. One example is Raupach (1984), who adopted a psycholinguistic perspective on the issue and defined formulae as planning units in language processing, the boundaries of which are marked by pauses, hesitation markers, and so on. He concluded that "at a lower level of proficiency learners display a great variety of idiosyncratic forms of planning behavior, especially

in their use of lexicalized fillers and modifiers" (Raupach, 1984, p. 134); they then gradually acquire the temporal patterning of the L2 as well as what Dechert has called "islands of reliability", idiomatic formulae and collocations. Another study relevant here is Yorio (1989), who examined the frequencies of conjugated and two-word verbs in 15 NS and 25 non-native speaker (NNS) college students' writing. He was astonished to see that

- the advanced learners used more prefabricated language than the beginners, which supports the idea that accurate idiomatic expression requires a certain degree of general language proficiency;
- the kinds of errors the learners made suggested that they did not treat these prefabs as fundamentally different from generated phrases, which undermines the distinction between lexis and grammar;
- differences between NS and non-native speaker (NNS) writers manifested themselves less in the *proportions* of two-word verbs used, but more in the *kinds* of verbs used, which again points to the difference between native-like fluency and native-like selection.

Implications of this phraseological perspective on L2 production accuracy for language teaching are discussed at length in Nattinger and DeCarrico (1992), who suggest the use of what they refer to as *lexical phrases*. Howarth (1998) presents a more fine-grained descriptive model of different kinds of constructions that was borrowed from Soviet phraseology research, distinguishing between free combinations, restricted collocations, and idioms. He points out the centrality of this theoretical concept for issues of accuracy in L2 production when he writes

[M]any learners fail to understand the existence of the central area of the phraseological spectrum between free combinations and idioms. It is in handling restricted collocations that errors of both a lexical and grammatical structure constantly occur. Moreover, learners need to understand that restricted collocations make up a significant part of a typical native speaker's production in both speech and writing. (Howarth 1998, p. 186)

The edited volume by Schmitt (2004) provides an overview of more recent research on the acquisition of formulaic language. Of particular relevance in this context is the contribution by Schmitt et al. (2004) on the results of a longitudinal study of EAP learners which suggest that relatively proficient EAP learners have a rich, and continuously growing, repertoire of formulaic sequences. Dörnyei et al. (2004), who investigated two learners, point out that three main factors seem to influence the acquisition of formulaic language: aptitude, motivation, and sociocultural adaptation. Supporting evidence for the latter comes also from Adolphs and Durow (2004), who present preliminary evidence that there is a positive correlation between successful acquisition of formulaic language and the degree of social integration of the learner in the target language environment. Finally, Spöttl and McCarthy (2004) present the first empirical

study of learners' knowledge of formulaic language across L1, L2, L3, and L4. Their results indicate that holistically processed phrases are typically available for inter-language transfer, and also confirm a positive correlation between formulaic language knowledge and general language proficiency.

Corpus studies of phraseology in L2 production

In corpus-linguistics, the idea of a side-by-side of rule-governed and schematic language has been a long-standing working hypothesis. Maybe the most striking corpus-linguistic description of this dual nature of language was given by Sinclair (1991), who referred to them as the *Open Choice* and the *Idiom Principle*, respectively. Accordingly, corpus-linguistic concepts like that of collocation, colligation (Firth, 1968, p. 181), semantic prosody (Sinclair, 1991), and even full-blown descriptive frameworks such as Hunston, and Francis' (2000) *Pattern Grammar* are based on the assumption that meaning always emerges contextually in the interplay of constructions (even if not every corpus linguist would use the term *construction*). It appears, then, that corpus linguistics is theoretically compatible with a definition of L2 accuracy as adequate selection; moreover, corpus data present a potential solution to the problem of data scarcity alluded to in recent studies such as Schmitt (2004).

However, it is only since the launch of learner corpora like the *International Corpus of Learners English* (ICLE) that corpus linguists have begun more systematically to investigate the implications of this assumption for descriptions of learner language, acquisition processes, and language teaching. The state of the art of corpus-linguistic phraseological research in language learning and teaching can be glimpsed from Meunier and Granger's (2008) edited volume. Handl (2008), for instance, sets out to "find a systematic procedure for selecting collocations from authentic language and displaying them in dictionaries aimed at non-native speakers of English" (p. 44). She presents a multi-dimensional profile for collocations (including lexical, semantic, and statistical information) and suggests ways to display this bundle of information in an accessible way. She points to the relevance of quantitative approaches to collocations: "[i]t is with the help of the collocational factor responsible for the statistical dimension that a systematic picture of the internal structure of collocations can be drawn" (Handl, 2008, p. 62).

Osborne (2008) examines the occurrence of typical errors of learners of English (including omission of 3rd person *-s*, inappropriate adverb placement, and plural use of mass nouns) and finds that they are partially motivated by contextual effects. The three major effects he identifies are blending, when items used together share or transfer their features (as in *drugs are an issue which arouse strong feelings*); bonding, when collocational links override syntactic requirements (e.g. *follow blindly everything*); and burying, when elements embedded in larger units become less salient and lose obligatory grammatical features (as in *He ... loves when a tender and careful woman waits for him ... and ... meet him with a kind smile*).

Another study in that volume is Paquot (2008), who considers "the potential influence of the mother tongue on learners' production of both correct and incorrect multi-word units that are typically used to fulfil an important rhetorical function, namely exemplification, in academic writing" (p. 101). She finds that multi-word expressions with a clearly delineated pragmatic function are more easily transferred from the L1, and that transfer of form usually also entails transfer of knowledge about the frequency and preferred register of the expression in question.

Corpus-based studies on constructions in L2 production

In an earlier study (Gries & Wulff, 2005), we combined corpus-based and experimental evidence to address the questions (i) whether argument structure constructions can be argued to be a part of second language learners' mental lexicon, and (ii) to what extent language learners are aware of the construction-specific verb preferences of these constructions (which were obtained from NS corpus data; cf. Gries & Stefanowitsch, 2004 and case study 1 below). To that end, we carried out a syntactic priming experiment (using a sentence completion task) and a semantic sorting experiment in which subjects could adopt either a verb-based or a construction-based sorting strategy. The experimental results were then correlated with corpus data from (i) the ICE-GB as an L1 corpus and (ii) verb-subcategorization preferences in a parsed L1 German corpus (cf. Schulte im Walde, 2006). In sum, the results showed that (i) learners do exhibit syntactic priming and semantic sorting preferences that strongly support the assumption that constructions are part of their interlanguage lexicon, and (ii) the priming effects closely resemble those of NS of English in that they are very highly correlated with NSs' verbal subcategorization preferences, but at the same time completely *un*correlated with the subcategorization preferences of the German translation equivalents of these verbs (ruling out simple transfer from L1).

In a follow-up study (Gries & Wulff, 2009), we examined whether similar evidence can be gathered for English constructions other than argument structure constructions. A corpus analysis of gerund and infinitival complement constructions from the British component of the *International Corpus of English* identified the verbs distinguishing best between these two constructions. These were used as experimental stimuli in a sentence-completion and a sentence- acceptability rating experiment. The results supported the hypothesis that gerund and infinitival complement constructions have attained some kind of constructional status for the L2 learners: both patterns exhibit verb-specific constructional preferences and priming effects.

A third study that is important to mention in the present context is Liang (2002), who replicated the sorting experiment with Chinese learners of English at different proficiency levels: with beginners, who had had two years of English instruction; intermediate learners, who had passed the national entrance exam to college; and with advanced learners, who had passed the Chinese national test for non-English majors.

Liang found that the more proficient learners increasingly relied on construction-based sorting. In this way, L2 learners are apparently very similar to children acquiring their first language in that constructional knowledge beyond the word level is gained over time, and therefore one indicator of general language proficiency. On the other hand and interestingly, the most advanced learner group – German learners of English with a median number of years of instruction of more than 11 years – relied more on the constructions than the native speakers in Bencini and Goldberg (2000). One way to explain this result involves the assumption that the learners notice the probabilistic patterning in English that ultimately gives rise to native speakers' sorting preferences (cf. also Ellis & Ferreira-Junior 2009), but then turn it into a more absolute pattern or maybe even a rule and apply it more rigorously and less flexibly than native speakers.[1]

A constructionist approach to accuracy in L2 production

Let us begin by summarizing the main conclusions from the review of literature:

- Accuracy cannot be defined (exclusively) as a rule-based, binary concept. Instead, a major component (if not the most important one) is native-like selection, a highly context-dependent and inherently scalar phenomenon.
- The growing awareness for the intricate interplay between constructions has changed our definition of prefabricated language. Rather than seeing prefabricated and rule-based language in opposition, we assume a continuum of differently schematized constructions.
- Learners display sensitivity to this continuum in various ways. Differences in all the various parameters characterizing this continuum (including semantic transparency, pragmatic function, and frequency) are good predictors of learners' relative difficulty with acquiring a given construction. This manifests itself also in the kinds of errors learners produce, which are often accountable by reference to contextual factors.
- For advanced learners, evidence has been provided that they have some knowledge even of highly schematic constructions and their interactions with other constructions that resembles that of NS (but may in fact be more rigid).
- The mastery of constructions and their systematic associations with other constructions is a gradual process. Idiomatic expression follows the acquisition of individual words, and (stock) phrases.

1. This pattern is again reminiscent of processes in first language acquisition where children are initially sometimes very rigid in their use of words and constructions and where their later acquisition involves a relaxation of what children perceived to be all-or-nothing rules into the more adult-like probabilistic pattern (cf. Stoll & Gries, 2008, for an example from the acquisition of Russian tense-aspect patterning).

Given all these findings, we suggest the following constructionist definition of accuracy:

(1) Accuracy in L2 production is the selection of a construction (in the Goldbergian sense of the term) in its preferred context within a particular target variety and genre.

The notion of context deserves some elaboration here. First of all, we intend the term to cover two meanings: it can mean that one construction occurs with another construction more often than with other, competing constructions. The most straightforward example for this would be a verb occurring more often in one syntactic frame than another (recall the example of *give* and the ditransitive above). At the same time, however, sensitivity to context can also manifest itself in linguistic features such that a construction will prefer to occur with certain elements of another construction. A well-known example for this form of selection is the preference of certain verbs to occur in the ditransitive construction particularly strongly if the subject noun phrase of the ditransitive construction is animate (again, *give* in the ditransitive is a case in point).

This definition of accuracy embraces the findings above in various ways. It does not rely on a strict separation of lexis and grammar, but shifts the focus on constructions in interaction and, especially given our operationalization proposed below, allows for an integration of lexical use (as argued for by Skehan, 2009). Given the definition of construction in Construction Grammar, our definition of accuracy is by no means restricted to the interaction of words and syntactic frames, as in Pawley and Syder's definition of native-like selection. Instead, this definition can also involve the morphological, syntactic, and pragmatic specifications of constructions. Similarly, our definition of context allows us to describe any systematic associations between constructions and their linguistic environment, down to features like animacy, constituent length, definiteness, information status, pragmatic function, or the like. Last but not least, our definition of accuracy reflects our understanding of language proficiency as a gradual phenomenon that transcends a NS-NNS divide.

Ultimately, a scientific definition is only as good as its potential to be tested and measured. As regards our definition of accuracy, its value crucially hinges on the notion of *construction* as a linguistic entity that can be clearly identified, as well as the notion of preferred context, which entails that not only do we have to be able to identify the context, we furthermore need to be able to distinguish preferred contexts from dispreferred ones, which we will do with a corpus-linguistic approach. The specific corpus-linguistic method that is perfectly compatible with our concept of accuracy is collostructional analysis'.

Measuring accuracy: Collostructional analysis

Collostructional analysis refers to a family of related corpus-linguistic methods developed by Gries and Stefanowitsch (Stefanowitsch & Gries, 2003; Gries &

Stefanowitsch, 2004), all of which measure the association between two constructions (as defined above). All these methods are text-internal lexical measures compatible with the definition of accuracy outlined above in (at least) two major regards: while typically applied to measuring the association between words and more complex constructions (such as the syntactic frames they occur in), collostructional analysis is not restricted to measuring association at the syntax-lexis interface, but can take as its input any two linguistic entities. (In fact, the method, unlike the definition, is not even restricted to measuring intra-constructional associations: it is perfectly feasible to use the same method to, say, measure different aspects of phonetic/phonological accuracy by looking into associations between phones, phones and morphemes, phones and words, etc.) Collostructional analysis is a technical operationalization of accuracy when defined as native-like selection, asking: what is the likelihood of a construction X in the environment of another construction Y?

We give a first idea of the wide applicability of collostructional analysis below by presenting the results of two different case studies in which patterning in the language of learners is compared to the, so to speak, baseline of patterning in the language of native speakers.[2] The first case study looks at associations between argument structure constructions and the matrix verbs that occur in them. The second case study examines the occurrence of matrix verbs depending on the morphological realization of a complement verb.[3] More precisely, both case studies consider the association between verbs and not just one other construction, but two variants of constructions, respectively: in case study one, we examine which verbs are specifically associated with one of two argument structure constructions that are often assumed to alternate more or less freely, the ditransitive and the prepositional dative. In case study two, we consider

2. Gilquin (to appear) actually makes a very similar point to the one we are trying to make here. She also demonstrates the usefulness of collostructional analysis for comparing the verbs associated with periphrastic causative constructions in NS and NNS data. Since causative constructions are relatively rare, Gilquin pooled ICLE data from 15 different L1 backgrounds. Her results show a rather poor fit between NS and NNS data, and she discusses lack of register awareness, transfer from L1, and inadequate teaching materials as potential factors responsible for this result. Two additional factors to be taken into consideration are the scarcity of her data and the pooling of so many different L1 backgrounds. Nevertheless, it is interesting to see that the fit between NS and NNS preferences is so much poorer for a relatively infrequent construction like causatives – from a usage-based perspective, we would actually predict this result. Further research on measuring language proficiency along dimensions of verb-construction associations in different frequency bands would be desirable to address this issue more systematically.

3. Note that the collostructional approach takes into consideration not just the mere frequency of co-occurrence of a word and a construction (or a word and a register), it also takes into consideration the overall frequencies of the word and the construction. In this regard, this method is superior to the raw-frequency approach by the otherwise very comprehensive Longman Grammar of Spoken and Written English (Biber et al. 1999). Other applications of collostructional analysis include studies of dialectal variation (Wulff, Gries, & Stefanowitsch, 2007; Mukherjee & Gries, 2009) and diachronic stages (Hilpert, 2006; Gries & Hilpert, 2008).

the preference of a given verb to occur with either gerundial or infinitival complements – another alternation that frequently features in L2 teaching materials.

In order to assess this distinctive association of a given verb with either of the two respective constructional choices, we employed one specific member of collostructional analysis, a so-called *Distinctive Collexeme Analysis* (DCA). Lexemes that are significantly associated with one construction as opposed to the other (that is, ditransitive vs. prepositional dative or gerundial vs. infinitival complementation, respectively) are referred to as distinctive collexemes of that construction. To test whether a given verb lemma is a distinctive collexeme of either argument structure or complementation construction, four frequencies are entered into a 2-by-2 table:

– the token frequency of that lemma in construction$_1$;
– the token frequency of that lemma in construction$_2$;
– the frequency of construction$_1$;
– the frequency of construction$_2$.

A Fisher-Yates exact text is applied to that table, providing a *p*-value which is, for ease of exposition, log-transformed to the base of ten and multiplied with –1 (cf. Stefanowitsch & Gries, 2003:217–8 for justification of using the Fisher Yates exact test; other association measures can of course also be applied, for example in cases where the objective is to quantify absolute strengths of attraction or to compare data from different sample sizes). Accordingly, any *p*-value 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. For both case studies, we first retrieved all relevant frequencies for all verb lemmas attested in the two argument structure and complementation constructions and then computed the DCA with Coll.analysis 3 (Gries, 2004). (Note in passing that the kind of data entering into a DCA can also form the basis to explore lexical variety, and thus productivity, in constructional slots.)

Case studies

Ditransitive and prepositional dative in L1 and L2 production

As we mentioned earlier, English allows the expression of transfer and (often metaphorically) related senses with two major syntactic patterns, or constructions: as a ditransitive construction as in (2), or as a prepositional dative construction as in (3).

(2) Stefan showed Pat the paper.
(3) Stefan showed the paper to Pat.

Cognitive-linguistic studies have carved out subtle, yet systematic meaning differences between the two constructions which become most transparent in the lexical semantics

of the verbs that preferably occur in either construction (cf. Goldberg, 1995, ch. 6).[4] Interestingly, corpus analyses in cognitive linguistics have shown that, in NS data, these meaning differences are strongly reflected in certain verbs being distinctively associated with either one of these constructions (cf. below and Gries & Stefanowitsch, 2004).

Let us look at such NS data first. Gries and Stefanowitsch (2004) extracted all verb lemmas occurring in the ditransitive and/or the prepositional dative construction from the British component of the *International Corpus of English* (ICE-GB). After manual cleaning of the data, they obtained 339 different verb lemmas occurring in either construction, totalling 2,954 verb tokens (1,035 in the ditransitive construction and 1,919 in the prepositional dative construction) and then ran a DCA. Table 1 displays the

Table 1. Collexemes distinguishing the ditransitive and prepositional dative constructions in NS English (ICE-GB) (from Gries & Stefanowitsch, 2004, p. 106)

Ditransitive		**Prepositional dative**	
Collexeme	**-log$_{10}$ *p***	**Collexeme**	**-log$_{10}$ *p***
give (461:146)	119.74	*bring* (7:82)	8.83
tell (128:2)	57.06	*play* (1:37)	5.84
show (49:15)	11.08	*take* (12:63)	3.74
offer (43:15)	9	*pass* (2:29)	3.65
cost (20:1)	8.01	*make* (3:23)	2.17
teach (15:1)	5.83	*sell* (1:14)	1.86
wish (9:1)	3.27	*do* (10:40)	1.82
ask (12:4)	2.89	*supply* (1:12)	1.54
promise (7:1)	2.45	*read* (1:10)	1.22
deny (8:3)	1.91	*hand* (5:21)	1.2
award (7:3)	1.59	*feed* (1:9)	1.07
grant (5:2)	1.26	*leave* (6:20)	0.86
cause (8:9)	0.67	*keep* (1:7)	0.77
drop (3:2)	0.63	*pay* (13:34)	0.74
charge (4:4)	0.53	*assign* (3:8)	0.37
get (20:32)	0.46	*set* (2:6)	0.37
allocate (4:5)	0.41	*write* (4:9)	0.3
send (64:113)	0.4	*cut* (2:5)	0.28
owe (6:9)	0.36	*lend* (7:13)	0.22
lose (2:3)	0.24		

4. These semantic differences, together with other distributional characteristics, strongly suggest treating each syntactic pattern as a construction in its own right rather than just as simple alternants (cf. Goldberg, 2002); our present discussion of these two constructions in terms of an alternation is purely a matter of terminological convenience and no theoretical significance should be attached to it.

25 verbs distinctively associated with either construction, in descending order of distinctiveness; the numbers in parentheses are the frequencies in the ditransitive and prepositional dative construction respectively (we report verbs that yielded a -log *p* value of 1.3 or higher, or that occur at least three times in either construction).

As Gries and Stefanowitsch (2004, p. 106–7) point out, *give* and most other distinctive collexemes of the ditransitive construction denote some form of transfer (literal or metaphorical) involving direct contact between an agent and a recipient. The distinctive collexemes for the prepositional dative construction, on the other hand, often involve some distance between the agent and the recipient that must be overcome to complete the transfer; that is, the patient is moved along some path to the recipient, which is why this construction is often referred to as the caused-motion construction. They also note that all verbs denoting commercial transactions are distinctive for the prepositional dative, with the exception of *cost*, which they attribute to the fact that this verb, unlike the other commercial transaction verbs, does not involve motion and thus better fits the semantics of the ditransitive. Moreover, they point out that looking at the verbs that do *not* yield the significance threshold of 1.3 can be revealing too: they identify *lend*, *send*, *get*, and *write* as the verbs alternating most freely between the two constructions.

Given these findings, the dative alternation makes for an interesting case study in an ESL context: are (advanced) learners also aware of these construction-specific verb preferences? If not, what kind of patterning, if any, do they exhibit? If yes, do they use verbs more or less flexibly than NS? As mentioned above, in Gries and Wulff (2005) we provided experimental evidence that the NNS data pattern similar to NS and, in the case of the sorting, were even more extremely construction-based than the NS. Here, we will use NNS corpus data, complementing Gries and Stefanowitsch's results with data from the German and Dutch sub-corpora of the *International Corpus of Learner English* (ICLE). An exhaustive retrieval and manual inspection yielded 34 different verb types and 623 tokens (450 for the ditransitive and 173 for the prepositional dative construction).[5] Table 2 summarizes, in analogy to Table 1 above, the results of the DCA for the advanced learners of English represented in ICLE.

Comparing Tables 1 and 2, we see that the overall results are indeed highly similar. Overall, the advanced learners seem to have recognized that the ditransitive construction preferably takes verbs denoting transfer with direct contact between agent and recipient; with regard to the most strongly associated collexemes distinctive for the ditransitive, the NS and the NNS lists are nearly identical (there is only some minor variation in the ranking). Looking at the most distinctive collexemes of the prepositional dative, however, we find some interesting deviations from NS use. First, *send* fits the semantics of the prepositional dative/caused-motion construction perfectly, but

5. The smaller total sample size was the reason why we pooled data from two different L1 backgrounds here (cf. also note 2). Note that this does not speak to the limitations of the method per se, but only to the limited availability of corpus data.

Table 2. Collexemes distinguishing the ditransitive and prepositional dative constructions in NNS English (D/G-ICLE)

Ditransitive		Prepositional dative	
Collexeme	**-log$_{10}$ *p***	**Collexeme**	**-log$_{10}$ *p***
give (268:56)	9.09	*grant* (8:2)	0.35
show (39:3)	3.17	*send* (1:28)	14.97
tell (26:1)	2.83	*pay* (3:20)	8.6
cost (11:0)	1.57	*bring* (10:20)	5.22
buy (7:0)	0.99	*write* (0:6)	3.37
teach (11:1)	0.96	*do* (2:7)	2.61
offer (24:5)	0.86	*deliver* (0:4)	2.24
ask (8:1)	0.63	*owe* (1:5)	2.13
assign (4:0)	0.57	*sell* (4:7)	1.88
guarantee (4:0)	0.57		

surprisingly still does not significantly prefer that construction in the NS data. On the other hand, in the NNS data we find the perfect match that one would have expected to see in the NS data: *send* is the strongest collexeme for the caused-motion construction. Again and just as in the sorting data, the NNS exhibit a behavior that is in fact more in the expected direction than that of the NS and illustrates learners' tendencies to form very strong generalizations.

A second interesting aspect of the results is that there are two kinds of verbs that prefer the caused-motion construction in the NNS data: verbs that prefer the same construction in the NS data (such as *bring*), and verbs that exhibit no strong preference for either construction in the NS data (such as *owe*, *write*, and *pay*). This may be because of a learner strategy to assign verbs they have not heard/seen being used predominantly in one pattern to the construction for which there is less of a translational equivalent in Dutch and German. However, when looking at transfer, we see that transfer from L1 can be misleading: in the NS data, *guarantee*, which does not even occur in the NS data list (likely because *grant* fills that semantic niche already) yields a significant value. Its presence can be accounted for by its frequent occurrence in German.

Irrespective of what is ultimately the main reason for these patterns, a distinctive collexeme analysis can help identify non-idiomatic choices of advanced learners both on the more general level (i.e., when different speakers are pooled, as in the above case) or on the more individual level (i.e., when we use its results to determine why a NNS has used a verb-construction combination that NS typically disfavor).

The overall good correlation between the NS and NNS preferences can be quantified in terms of a correlation: Kendall's *tau* = 0.7; $z = 5.46$; $p < 0.001$. Figure 1 provides a graphical representation of this correlation (including only significantly distinctive collexemes occurring in both the NS and NNS corpora). In order to avoid scaling

issues, the -log *p* values were normalized between –1 and +1 by setting the smallest value obtained from either data set to –1, the maximum value to +1, and assigning all values in between a normalized frequency that reflects their distance from these two extremes. Values around 0 mean that the verb has no preference for either construction; values higher than 0 mean that the verb is positively associated with the ditransitive construction; and values lower than zero mean that the verb is negatively associated with the ditransitive construction, or, in other words, positively associated with the prepositional dative construction. The numbers at the ±0.5/±0.5 data points in the grid provide us with a more general summary of the results: the 17 in the top right corner, for instance, means that 17 verbs have the same (positive) attraction to the ditransitive in the NS and the NNS data; 6 verbs have the same (negative) attraction; one verb is positively associated with the ditransitive in the NS data, but significant for the prepositional dative in the NNS data; and for one other verb, it is exactly the other way around. So in sum, for 24/26 verbs, we see a match between the verb-specific constructional preferences between NS and NNS – a result that again underscores how well the NNS have extracted the distributional patterns in their L2 language input.

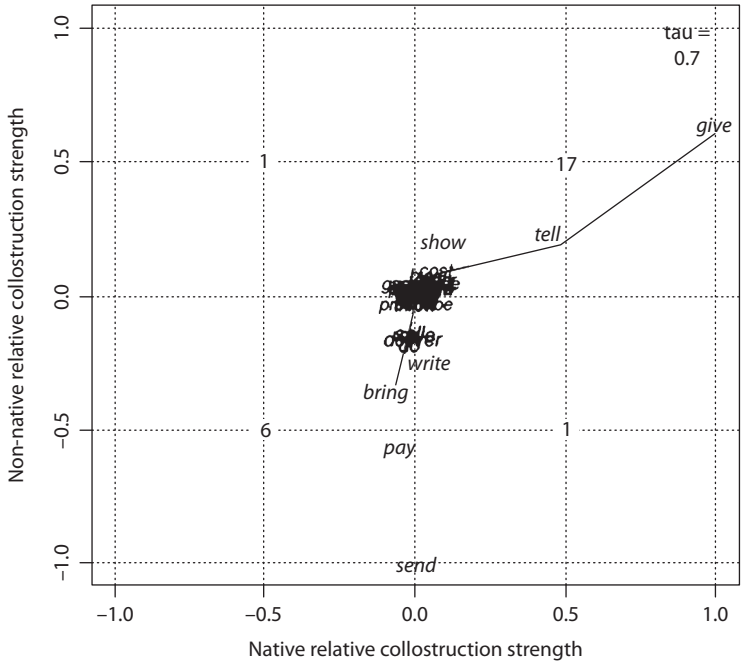

Figure 1. Correlation between NS and NNS relative collostruction strengths between verbs and the ditransitive construction

Infinitival and gerundial complementation in L1 and L2 production

In our second case study, we look at another pair of constructions, infinitival and gerundial complementation constructions; examples are given in (4) and (5), respectively.

(4) Steffi began to feed the squirrels.
(5) Steffi began feeding the squirrels.

These two constructions were shown to present difficulties even to advanced learners of English (cf. Celce-Murcia & Larsen Freeman, 1999, p. 645; Schwartz & Causarano, 2007). This may have to do with the fact that the semantics of the constructions are arguably much less tangible than in the dative alternation, where both alternatives encode perceivable and readily interpretable universal humanly relevant scenes (as opposed to the less tangible aspectual meanings of the two complementation constructions). Another factor that clearly plays a role is that equivalents of the infinitival complementation construction are much more prominent cross-linguistically, enabling positive transfer, while the gerundial complementation construction is comparatively rare, and in languages that have both constructions, the infinitival complementation construction tends to be considerably more frequent (cf. Butyoi, 1977; Mair, 2003).

As with the first case study, let us first look at the NS data. Table 3 provides a summary of the data obtained by Gries and Wulff (2009) from the ICE-GB; the data set comprised 480 tokens of the gerundial complementation construction (48 different verb types) and 2,863 tokens of the infinitival complementation construction (98 different verb types), totaling 120 verb types overall.

Looking at Table 3, we see some established claims about the semantic differences between the two constructions confirmed. For one, the verbs most distinctively associated with the infinitival construction, *try* and *wish*, both denote potentiality, while the verbs most distinctive for the gerundial construction, *keep*, *start*, and *stop*, denote actual events. Along similar lines, many of the collexemes distinctive for the infinitival construction are future-oriented (*intend*, *hope*, *learn*, and *aim* are just a few examples here), while the distinctive collexemes of the gerundial construction evoke an interpretation in relation to the time of speaking (for example *avoid*, *end*, *imagine*, *hate*). Interestingly, for *begin*, which is often featured in teaching materials as being tied to the infinitival construction, and contrasted with the near-synonymous *start*, which is claimed to prefer the gerundial construction, the corpus data provide a much less rigorous picture: *start* is indeed highly distinctive for the gerundial construction, but *begin* is far from being significantly associated with the infinitive – on the contrary, the DCA, which takes not only the raw frequencies of occurrence, but also the general frequency of *begin* in all its contexts into consideration, suggests a weak association with the gerundial construction. This example nicely illustrated how corpus linguistics may help improve instructional materials considerably by taking authentic data into consideration.

Table 3. Collexemes distinguishing the infinitival and gerundial complementation constructions in NS English (ICE-GB) (from Gries & Wulff, 2009)

Infinitival complementation		Gerundial complementation	
Collexeme	**-log$_{10}$ *p***	**Collexeme**	**-log$_{10}$ *p***
try (452:8)	22.44	*pretend* (10:0)	0.67
wish (79:0)	5.39	*keep* (0:87)	76.45
manage (70:0)	4.77	*start* (89:96)	35.23
seek (64:0)	4.35	*stop* (4:40)	29.45
tend (123:5)	4.06	*avoid* (0:14)	11.87
intend (54:0)	3.67	*end* (0:14)	11.87
attempt (47:0)	3.19	*enjoy* (0:14)	11.87
hope (47:0)	3.19	*mind* (0:14)	11.87
fail (60:1)	3.09	*remember* (10:20)	10.14
like (208:17)	3.03	*go* (31:26)	7.99
refuse (44:0)	2.98	*consider* (15:15)	5.45
learn (31:0)	2.1	*envisage* (0:4)	3.38
plan (28:0)	1.89	*finish* (0:4)	3.38
continue (103:9)	1.53	*carry* (0:3)	2.53
afford (22:0)	1.49	*fancy* (0:3)	2.53
force (18:0)	1.22	*imagine* (0:3)	2.53
prefer (18:0)	1.22	*resist* (0:3)	2.53
aim (17:0)	1.15	*catch* (0:2)	1.69
tempt (14:0)	0.94	*hate* (3:3)	1.38
encourage (13:0)	0.88	*bear* (1:2)	1.25
claim (11:0)	0.74	*begin* (119:27)	1.03
forget (11:0)	0.74	*recommend* (2:2)	0.99

Again, we complement the NS with NNS data. For this case study, we could restrict our search to the German component of ICLE since this gave us a sufficient number of hits already. An exhaustive retrieval resulted in 72 verb types and 899 verb tokens overall (230 for the gerundial construction, 669 for the infinitival construction after manual inspection for false hits). Table 4 displays the results of the DCA for these data (again, we display all collexemes that either yielded a -log *p* value of 1.3 or higher, or that occur at least three times in either construction).

Comparing Tables 3 and 4, we see that there are many commonalities, but the match between the NS and the NNS data is not as good as in our first case study, which is probably due to the less tangible constructional semantics of the two target constructions. As far as the most distinctive collexemes are concerned, the match is very good again: *try*, *manage*, *like*, and *tend* range among the collexemes most distinctive for the infinitival construction; *keep*, *go*, *stop*, *start*, *avoid*, and *enjoy* occupy the top

Table 4. Collexemes distinguishing the infinitival and gerundial complementation constructions in NNS English (G-ICLE)

Infinitival complementation		Gerundial complementation	
Collexeme	**-log$_{10}$ *p***	**Collexeme**	**-log$_{10}$ *p***
try (256:0)	39.9	*fail* (6:1)	0.37
manage (38:0)	5	*hope* (5:1)	0.28
like (72:6)	4.54	*keep* (0:23)	13.99
tend (28:0)	3.66	*go* (4:29)	13.6
learn (26:1)	2.5	*stop* (2:19)	9.4
begin (25:1)	2.38	*start* (54:55)	8.71
dare (23:2)	1.58	*avoid* (1:12)	6.2
forget (10:0)	1.29	*enjoy* (1:12)	6.2
wish (10:0)	1.29	*end up* (0:6)	3.57
refuse (6:0)	0.77	*give up* (0:4)	2.38
attempt (4:0)	0.51	*continue* (1:5)	2.3
promise (4:0)	0.51	*hate* (1:5)	2.3
intend (3:0)	0.39	*remember* (1:5)	2.3
strive (3:0)	0.39	*finish* (0:3)	1.78
succeed (3:0)	0.39	*keep on* (0:3)	1.78
unlearn (3:0)	0.39	*go on* (1:4)	1.78
afford (6:1)	0.37	*prefer* (9:8)	1.36

ranks in the gerundial collexeme list, which testifies to the learners' ability to accurately select the idiomatic complementation construction for these verbs. But some selections stand out as clearly not native-like. *Prefer* and *continue*, for instance, are significantly associated with the gerundial construction in the NNS data but attracted to the infinitival construction in the NS data. Also, in accordance with teaching materials but in contrast to real NS usage, *begin* is strongly preferred in the infinitival construction. Similarly, *fail* and *hope* do not nearly rank as high in the infinitival construction collexeme list in the NNS data as they do in the NS data. Maybe most striking is the German learners' overuse of phrasal verbs such as *end up*, *give up*, *keep on*, and *go on* in the gerundial complementation construction. Note how all these verbs have the proper time reference and denote actuality, so they do fit the semantic constraints of the gerundial construction; in that sense, they are good examples of the intricacies of native-like selection that even advanced learners of English face. As German NSs ourselves, we can only speculate what the underlying motivation for the frequent use of these verbs may be. One possibility may be an attempt to transfer a very common construction in German *X ist am* $V_{infinitive}$ (*X is* V*ing*): the combination of the preposition *am* with the bare form of a verb is one of the more typical ways to express progressive aspect in German. The semantics of the gerundial complementation

construction are sufficiently compatible with a progressive reading, and learners may fill the slot of the German *am* with the particle of the phrasal verb.[6]

On a final note, a comparison of the NS and the NNS data also helps us to identify several verbs that do not figure in the learner data at all and are therefore primary candidates for further teaching: *seek* and *continue* are two example of verbs distinctively associated with the infinitival construction; *envisage*, *fancy*, and *imagine* are but three examples of verbs distinctively associated with the gerundial construction that do not appear in the NNS data at all.

The overall slightly less impressive correlation (compared to the first case study) is also obvious in the graphical display in Figure 2 (Kendall's *tau* = 0.61; $z = 5.71$; $p < 0.001$). Looking at the number at the ±0.5/±0.5 grid points again, we find that while the majority of verbs (14 + 15 = 29) are associated with the same construction in both the NS and the NNS data, there are six verbs (3 + 3) that are distinctive for one construction in the NS data, but distinctively associated by the NNS with the other construction, and vice versa.

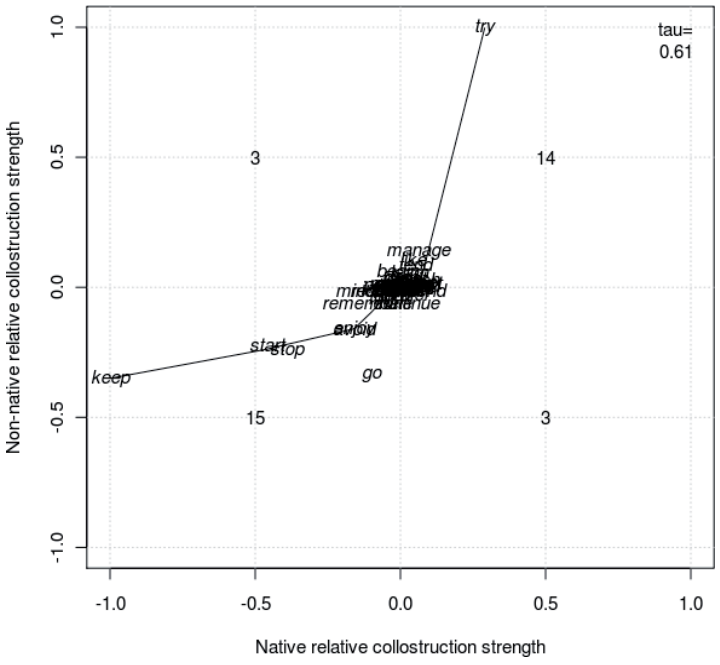

Figure 2. Correlation between NS and NNS relative collostruction strengths between verbs and the infinitival complementation construction

6. As one anonymous reviewer pointed out, two other possible motivations for this overuse of phrasal verbs by German learners are that phrasal verbs feature very prominently in learner text books, and that learners may transfer the high frequency of phrasal verbs in spoken language (cf. Biber et al. 1999: Section 5.3.2) to their written essays (on learners' tendency to be driven in writing by their oral language proficiency, see Gilquin & Paquot, 2008).

Discussion

Both the theoretical perspective adopted here in general and the definition of accuracy proposed above in particular have several implications for instruction and task design. In this section, we discuss a few of these implications and relate them to currently widely-discussed topics in the SLA community. In the following section, we first briefly discuss the question of how, from our perspective, learners become more accurate over time, before we then turn to instructional design.

How learners' production becomes more accurate

Our theoretical affinity to the framework of Construction Grammar and our definition of accuracy are obviously closely related to approaches in usage-based cognitive linguistics as well as exemplar-based connectionist models in psycholinguistics. Learning – i.e., among other things, becoming more accurate – involves an intuitive data-driven statistical learning process of learners

- noticing forms f_1, f_2, ... that instantiate patterns p_1, p_2, ... and serve functions x_1, x_2, ... in the input;
- storing either the specific exemplars f_1, f_2, ... or more schematic generalizations of them in a complex multi-dimensional space, whose dimensions involve phonological, morphological, syntactic, semantic, pragmatic, register, and other distributionally or functionally noticeable dimensions;
- gradually fine-tuning this multi-dimensional space through the addition of additional exemplars or schemas so that emerging scatterclouds give rise to constructions (of various levels of granularity).

More succinctly

> [...] acquisition depends on exemplar learning and retention, out of which permanent abstract schemas gradually emerge and are immanent across the summed similarity of exemplar collections. These schemas are graded in strength depending on the number of exemplars and the degree to which semantic similarity is reinforced by phonological, lexical, and distributional similarity.
>
> (Abbot-Smith & Tomasello, 2006, p. 275)

(Cf. Ellis, 2007, for discussion of the Associative-Cognitive CREED for a more comprehensive overview). Thus, accuracy will increase proportionally to the extent that learners succeed in making the right generalizations regarding which form (e.g., the ditransitive or the caused-motion construction) is mapped onto which function (e.g., referring to the direct transfer of a concrete object from one human to another). Note that "making the right generalizations" amounts to nothing else than learners being able to extract prior probabilities (e.g., the knowledge that *give* is more frequent than *donate*) as well as posterior/conditional probabilities (e.g., the knowledge that

give is used ditransitively more often than *donate*) from the multidimensional input/space. The definition of accuracy proposed above not only explicitly incorporates such a probabilistic approach but is therefore also compatible with current theories of language production and, as a measure of co-occurrence strength, also easily extendable to handle the kind of multidimensional approaches to syntactic complexity argued for by Norris and Ortega (2009).

The ability to make the right generalizations about such form-function mappings in turn depends on a variety of individual learner characteristics, but less idiosyncratically also on

- the frequency of particular form-function mappings in the input – function again understood broadly as including animacy, definiteness, length, etc.;
- the amount of attention/processing allocated to such mappings (which in turn is dependent on the complexity and interactivity of the task in which a form-function mapping is to be used); and
- the degree to which particular form-function mappings are recognizable, salient, relevant, and reliable.

According to our broad definitions of *context* and *function*, if a learner uses *give* in the prepositional dative construction (which is generally the dispreferred choice), then this would lower his accuracy score unless, for instance, the recipient NP is very long, in which case even native speakers would also use the prepositional dative. Crucially, the above is based on generalizations of verb/construction use across speakers and cases/contexts. However, this also entails that the necessary next analytical step involves an additional more fine-grained analysis, which is why we are now exploring how well we can predict NNS constructional choices on a case-by-case basis, i.e., in the tradition of research on syntactic alternations in theoretical and usage-based linguistics. This will allow us to determine whether NNS not only exhibit overall similar tendencies to NS, but whether their choices are also governed by the same factors to the same degrees.

The view of learning and accuracy we articulated above has implications for the design of instruction, materials, and tasks, to which we now turn.

Implications for task design

Given many corpus linguists' claims, it would seem as if the recommendations for instructional (task) design were straightforward: include as much naturalistic corpus data as possible so that the learners' pattern-matching abilities kick in and extract relevant patterns. However, the situation is not as straightforward as has often been assumed. While corpus linguists have in fact argued in favor of more naturalistic data in instruction and instructional materials, more often than not such demands were not backed up by empirical studies that demonstrated the superiority of such materials. It seems intuitively obvious that authentic data are better, but they are typically also

much noisier and, thus, likely to contain potentially conflicting cues for form-function mappings that make it harder for learners to arrive at the right generalization(s) – carefully-constructed examples or minimal pairs, on the other hand, are by definition not natural, but may be more successful at providing the learner with the right cues, and only the right cues. In the meantime, however, research from the Associative--Cognitive CREED (on both first and second language acquisition) has provided different kinds of results that bear on this issue with regard to:

- the design of instructional materials: we now know that the use of authentic expressions in teaching materials may be at odds with their use in authentic settings, and the distribution of these expressions in learner data may be correlated more with the former than the latter (cf. the use of *begin* above);
- the frequency of stimuli: we now know that increased frequency of exposure will overall increase the likelihood that a particular structure will be noticed, processed in more detail, and integrated into the learners' L2 network. Increased input frequency was shown to yield best results when exposure was distributed over time as opposed to short-term mass exposure (cf. Ambridge et al., 2006);
- the complexity of the task and the stimuli: we now know that authentic examples, even if they are more complex to process, are not automatically worse since higher task complexity may in fact result in more elaborate processing of the material by the learner (cf. Robinson's Cognition Hypothesis; cf. Robinson, 2003, p. 651; Robinson & Gilabert, 2007, p. 162). On the other hand, if the form-function mapping to be learned is too complex (cf. the Multidimensional Model or Processability Theory) or embedded in a noisy context full of conflicting cues, then it may not be noticed by the learner. Thus, two kinds of things are particularly necessary. First, we need (more) precise and more multidimensional measures of linguistic complexity on various levels of analysis. With regard to syntactic measures, traditional measures such as MLUs, average syntactic depths, IPSyn etc. are often useful approximations, but the kind of multivariate measures employed in corpus-linguistic register studies (in particular Biber's (1988) multidimensional approach or the various indices integrated into Coh-Metrix at <http://cohmetrix.memphis.edu/>) may do more justice to the intricacies of syntactic complexity. With regard to lexical complexity, we need more careful analysis of what constitutes lexical diversity (cf. Skehan, 2009 for discussion of TTR, D, lambda, and other measures). Second, we need measures that integrate syntactic and lexical complexity and variability, and the collostructional approach or similar approaches based on co-occurrence data may be useful, especially once speaker-specific analyses are added. Finally, we need a careful sequencing of instructional modules in accordance with learners' zones of proximal development (cf. Schmidt, 1990; Robinson, this volume; Robinson & Gilabert, 2007 and below);
- the noticeability of the form-function mapping: we now know that not only must the learner notice the form-function mapping in question, but the degree to which

this is possible interacts with complexity such that, in situations of impending cognitive overload, learners tend to focus on matters of meaning and would therefore benefit from being alerted to matters of form.

All these findings present a positive outlook on the use of corpus data in instruction. The primary goal of the present paper was to provide general examples of how corpus-linguistic methods like collostructional analysis can be employed to guide the selection of relevant input data. However, in order to provide language teachers with more concrete suggestions for the implementation of second language research into their teaching, more systematic studies of learners at different levels of language proficiency and from different L1 backgrounds are called for (cf. Seidlhofer's (2002) learning-driven paradigm). Unfortunately, while there are now many L1 corpora available for many languages or which can be constructed on the fly, the situation is much more dire for L2 corpora, and few resources other than the ICLE corpus, which comprises more than 3 million words of learner essays by advanced learners of English from 21 different L1 backgrounds, are available and widely used. This severely limits the kinds of questions that can be addressed, particularly with regard to constructionist research, which requires larger amounts of data. Given the current state of data and methodology, we therefore consider the compilation of more and larger learner corpora as well as the exploration of corpus-linguistically motivated complexity and accuracy measures as the prime ways in which corpus linguists should contribute to SLA research.

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review, 23,* 275–290.

Ambridge, B., Theakston, A., Lieven, E. M. V., & Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development, 21*, 174–193.

Bencini, G., & Goldberg, A. E. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language, 43*, 640–651.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Butoyi, C. A. (1977). *The accuracy order of sentential complements by ESL learners*. Unpublished M.A. thesis, UCLA.

Byrd, P. (2005). Instructed grammar. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 545–561). London: Routledge.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: an ESL/EFL teacher's course*. Boston, MA: Heinle and Heinle.

Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages*. London: Academic Press.

Ellis, N. C. (2007). The Associative-Cognitive CREED. In B. VanPatten & J. Williams (Eds.), *Theories of second language acquisition: an introduction* (pp. 77–95). Mahwah, NJ: Lawrence Erlbaum Associates.

Ellis, N. C., & F. Ferreira-Junior. (2009). Constructions and their acquisition: islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics, 7*, 187–220.

Ellis, R., (1984). *Classroom second language development*. Oxford: Pergamon.

Ellis, R., & Yan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*, 59–84.

Firth, J. R. (1968). A synopsis of linguistic theory. In F. R. Palmer (Ed.), *Selected papers of J. R. Firth, 1952–59* (pp. 168–205). London: Longman.

Gilquin, G. (to appear). Lexical infelicity in causative constructions: comparing native and learner constructions. In J. Leino & R. von Waldenfels (Eds.), *Analytical causatives*. Munich: Lincom.

Gilquin, G., & Paquot, M. (2008). Too chatty: learner academic writing and register variation. *English Text Construction,1*, 41–61.

Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2002). Surface generalizations: an alternative to alternations. *Cognitive Linguistics, 13*, 327–356.

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

Gries, St. Th. (2004). *Coll.analysis 3*. *R*-script. Available at http://tinyurl.com/collostructions

Gries, St. Th., & Hilpert, M. (2008). The identification of stages in diachronic data: variability-based neighbor-clustering. *Corpora, 3*, 59–81.

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

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

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24*, 287–297.

Hakuta, K. (1976). A case study of a Japanese child learning English. *Language Learning, 26*, 321–351.

Hanania, E. A. S., & Gradman, H. L. (1977). Acquisition of English structures: a case study of an adult native speaker of Arabic in an English-speaking environment. *Language Learning, 27*, 75–91.

Handl, S. (2008). Essential collocations for learners of English. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 43–66). Amsterdam: John Benjamins.

Hatch, E. (1972). Some studies in language learning. *UCLA Working Papers in Teaching English as a Second Language, 6*, 29–36.

Hilpert, M. (2006). Distinctive collexeme analysis and diachrony. *Corpus Linguistics and Linguistic Theory, 2*, 243–257.

Housen, A., & Kuiken, F. (2009). Complexity, fluency, and accuracy in second language acquisition. *Applied Linguistics, 30*, 461–473.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology* (pp. 161–186). Oxford: Clarendon.

Krashen, S., & Scarcella, R. C. (1978). On routines and patterns in language acquisition and performance. *Language Learning, 28*, 283–300.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*, 590–619.

Liang, J. (2002). *Sentence comprehension by Chinese Learners of English: verb centered or construction-based*. Unpublished M.A. thesis, Guangdong University of Foreign Studies.

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. 347–377). Berlin: Mouton de Gruyter.

Meunier, F., & Granger, S. (Eds.). (2008). *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins.

Mukherjee, J., & Gries, St. Th. (2009). Verb-construction associations in the International Corpus of English. *English World-Wide, 30*, 27–51.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics, 30*, 555–578.

Osborne, J. (2008). Phraseology effects as a trigger for errors in L2 English. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 67–83). Amsterdam: John Benjamins.

Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo: Kaitakusha.

Paquot, M. (2008). Exemplification in learner writing: a cross-linguistic perspective. In F. Meunier, & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 101–119). Amsterdam: John Benjamins.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: native-like selection and native-like fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). London: Longman.

Raupach, M. (1984). Formulae in second language speech production. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 114–137). Tübingen: Günter Narr.

Richards, J. C. (2002). Accuracy and fluency revisited. In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 35–52). Mahwah, NJ: Lawrence Erlbaum Associates.

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631–678). Malden, MA: Blackwell.

Robinson, P., & Gilabert, R. (2007). Task complexity, the Cognition Hypothesis and second language learning and performance. *International Review of Applied Linguistics, 45*, 161–176.

Scarcella, R. C. (1979). Watch up: A study of verbal routines in adult second language performance. *Working Papers on Bilingualism, 19*, 79–88.

Schmidt, R. W. (1983). Interaction, acculturation, and the acquisition of communicative competence: a case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137–174). Rowley, MA: Newbury House.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.

Schmitt, N. (Ed.). (2004). *Formulaic sequences*. Amsterdam: John Benjamins.
Schmitt, N., Dörnyei, Z., Adolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences: a longitudinal study. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 55–86). Amsterdam: John Benjamins.
Schumann, J. H. (1978). Second language acquisition: the pidginization hypothesis. In E. M. Hatch (Ed.), *Second language acquisition: a book of readings* (pp. 256–271). Rowley, MA: Newbury House.
Schwartz, M., & Lin Causarano, P. (2007). The role of frequency in SLA: an analysis of gerunds and infinitives in ESL written discourse. *Arizona Working Papers in SLA and Teaching, 14*, 43–57.
Seidlhofer, B. (2002). Pedagogy and local learner corpora: working with learning-driven data. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 213–234). Amsterdam: John Benjamins.
Shapira, R. G. (1978). The non-learning of English: case study of an adult. In E. M. Hatch (Ed.), *Second language acquisition: a book of readings* (pp. 246–255). Rowley, MA: Newbury House.
Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*, 510–532.
Spöttl, C., & McCarthy, M. (2004). Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 191–225). Amsterdam: John Benjamins.
Stefanowitsch, A., & Gries, St. Th. (2003). Collostructions: investigating the interaction between words and constructions. *International Journal of Corpus Linguistics, 8*, 209–243.
Stoll, S., & Gries, St. Th. (2008). How to characterize development in corpora: an association strength approach. *Journal of Child Language, 46*, 1075–1090.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Wagner-Gough, J. (1975). *Comparative studies in second language learning*. M.A. Thesis, UCLA, TESL Department.
Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics, 16*, 180–205.
Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawaii Press.
Wray, A.(2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
Wulff, S., Gries, St. Th., & Stefanowitsch, A. (2007). Brutal Brits and persuasive Americans: Variety-specific meaning construction in the *into*-causative. In G. Radden, K.-M. Köpcke, Th. Berg, & P. Siemund (Eds.), *Aspects of meaning construction in lexicon and grammar* (pp. 265–281). Amsterdam: John Benjamins.
Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss* (pp. 55–72). Cambridge: Cambridge University Press.