Cognitive Corpus Linguistics: five points of debate on current theory and methodology

Antti Arppe,[1] Gaëtanelle Gilquin,[2] Dylan Glynn,[3]
Martin Hilpert[4] and Arne Zeschel[5]

Abstract

Within cognitive linguistics, there is an increasing awareness that the study of linguistic phenomena needs to be grounded in usage. Ideally, research in cognitive linguistics should be based on authentic language use, its results should be replicable, and its claims falsifiable. Consequently, more and more studies now turn to corpora as a source of data. While corpus-based methodologies have increased in sophistication, the use of corpus data is also associated with a number of unresolved problems. The study of cognition through off-line linguistic data is, arguably, indirect, even if such data fulfils desirable qualities such as being natural, representative and plentiful. Several topics in this context stand out as particularly pressing issues. This discussion note addresses (1) converging evidence from corpora and experimentation, (2) whether corpora mirror psychological reality, (3) the theoretical value of corpus linguistic studies of 'alternations', (4) the relation of corpus linguistics and grammaticality judgments, and, lastly, (5) the nature of explanations in cognitive corpus linguistics. We do not claim to resolve these issues nor to cover all possible angles; instead, we strongly encourage reactions and further discussion.

[1] Department of General Linguistics, P.O. Box 24 (Unioninkatu 40), 00014 University of Helsinki, Finland.

[2] Centre for English Corpus Linguistics, Collège Erasme, Place Blaise Pascal 1, B–1348 Louvain-la-Neuve, Belgium.

[3] Språk- och litteraturcentrum, Helgonabacken 12, Box 201, 221 00 Lund, Sweden.

[4] Freiburg Institute for Advanced Studies (FRIAS), Albertstrasse 19, 79104 Freiburg, Germany.
 Correspondence to: Martin Hilpert, *e-mail:* martin.hilpert@frias.uni-freiburg.de

[5] ITALK, Institute of Business Communication and Information Science, University of Southern Denmark, Alsion 2, DK–6400 Sønderborg, Denmark.

1. Introduction

Within cognitive linguistics, there is an increasing awareness that the study of linguistic phenomena needs to be grounded in usage. Ideally, research in cognitive linguistics should be based on authentic language use, its results should be replicable, and its claims falsifiable. Consequently, more and more studies now turn to corpora as a source of data in what one might call Cognitive Corpus Linguistics. We use this term to refer to research that formulates questions about human cognition in such a way that statistical analysis based on corpus data can yield answers to these questions. As we see it, Cognitive Corpus Linguistics is aligned theoretically with other cognitive approaches to language. Shared issues include the commitment to engage in an interdisciplinary dialogue with the other cognitive sciences, the assumption that language structure emerges from language use, and the hypothesis that grammatical knowledge is non-modular. The last point entails that there is a lexico–syntactic continuum, and that the study of syntactic patterns falls under the purview of a cognitively orientated Construction Grammar (Langacker, 1987; and Goldberg, 2006).

While cognitive corpus linguistics has developed a range of sophisticated analytical methods, the use of corpus data is also associated with a number of unresolved problems. The study of cognition through off-line linguistic data is, arguably, indirect, even if such data fulfils desirable qualities such as being natural, representative, and plentiful. Still, the field of cognitively orientated corpus linguistics is growing (see Gries and Stefanowitsch, 2006; and Stefanowitsch and Gries, 2006) and, at the same time, becoming self-aware (Tummers *et al*., 2005; and Grondelaers *et al*., 2007). Meta-discussions on cognitive corpus linguistics often touch on the issue of converging evidence, and, hence, the interaction of corpora with other sources of data. Naturally, integrating different types of evidence is essential for cognitive linguistics and indeed usage-based linguistics at large. A view representing the current consensus is expressed by Tummers *et al*. (2005: 233):

> Overall, we will have to think of the empirical methodology of usage-based linguistics as having a helix-like structure, involving a gradual refinement of interpretations through a repeated confrontation with empirical data – all kinds of empirical data. An initial hypothesis, which may be derived introspectively, is confronted with the corpus data; interpreting the results leads to a more refined hypothesis and more questions, which may then be subjected to further experimental testing or a new confrontation with the corpus data – and so on.

In the pursuit of converging evidence, a growing number of studies now combine corpus-based methodologies with experimental designs. However, Gilquin and Gries (2009) point out that there is relatively little unity in the approaches that attempt to unite cognitive corpus linguistics and cognitive

psycholinguistics: corpus linguists make use of experimentation in a different way from classically trained psycholinguists, who in turn use corpora in a different way from that which is common in corpus linguistics. If converging evidence is to live up to its potential, more communication between the fields is needed.

In order to promote a critical discussion of these issues from within the discipline of corpus linguistics, the Freiburg Institute for Advanced Studies (FRIAS) invited a panel of cognitively orientated corpus linguists in October 2008. This paper is the result of the participants' collaborative effort to summarise the main ideas that were debated at the workshop. Like the workshop itself, the sections of this paper take a dialogical shape. Each contributor lays out an argument in an opening sequence while another participant offers a critical response to that argument. A final section ties together some loose ends – though by no means all of them. We hope that this format allows a more or less direct insight into the issues that the authors regard as important for the field today. We do not claim to resolve these issues; instead we strongly encourage reactions and further discussion.

2. Converging evidence

2.1 The case for converging evidence

> *Antti Arppe*:
> The benefits of multi-methodological research outweigh the problems
> – in linguistics as much as elsewhere.

Linguistics has suffered from methodological monism, with different schools having traditionally resorted to only one type of evidence and associated methods (see Hacking, 1996: 65–6; and Chafe, 1992: 92). Not surprisingly, researchers tend to look favourably on their own methods and data types while being somewhat sceptical towards the rest. However, each of these types of evidence and methods has its own particular restrictions, as discussed by Schütze (1996) with respect to introspection or Sampson (2007) with respect to corpora. Even more importantly, keeping to only one type of evidence or raising one evidence type above the others as more fundamental unnecessarily constrains the way we see language as an object of scientific study. In striving for a comprehensive understanding of language, we have to combine evidence types and methods that address different aspects of linguistic behaviour and knowledge.

Language is a multi-faceted phenomenon that has physiological, psychological, cognitive and social dimensions. As corpus linguists, we try to understand this multitude of interacting characteristics by studying aggregate data that pools the productions of many speakers and writers – often across different media, genres, registers, and even across different time periods. While richly annotated corpus data are thus well-suited to investigate

linguistic variation that is conditioned by structural, social, or textual factors, questions pertaining to the (linguistic component of) cognition of individual speakers need to be carefully operationalised if we want to analyse them on the basis of off-line data from corpora. Teasing apart the ontologically different factors governing the usage of a linguistic form is far from trivial, but we can point to successes. For instance, corpus-based work in cognitive sociolinguistics (Grondelaers *et al.*, 2008; Heylen *et al.*, 2008a; and Gries and Hilpert, 2009) illustrates how language-internal and language-external variables can be studied in an integrated fashion, thus allowing an assessment of their respective roles.

At the same time, it has to be conceded that some questions about the cognitive structure and processing of language are easier to study if an experimental approach is followed. An overarching consensus expressed by Haspelmath (2009) is that all types of data have their place and that no single investigative method can adequately cover all types of research questions relevant to language and its use. The investigative methods and types of evidence that are selected ought to follow from the characteristics of the phenomenon that is under investigation. These different considerations do not mean that linguists should settle for only one type of data and one associated type of research question. The reverse is true: they should strive to bridge the gaps that characterise the current research landscape. At the moment, there is an over-abundance of data types and methods available, and, as yet, no established understanding of their relationships.

One major challenge in multi-methodological research is how to ensure comparability of the operationalisations of hypotheses and research questions, as well as associated concepts, across different studies. In the study of *synonymy* and *semantic similarity*, we may hypothesise that near-synonymous words are differentiated by some systematic (but secondary) differences in their observable contexts of usage. In this, we may operationalise such context as the syntactic components of the argument structure, and the morphological and semantic characteristics of these components, which are linked with the same sentence to the near-synonyms under study. We may decide to focus either on a particular, restricted context, (e.g., finite forms of Russian *try* verbs preceding an infinitive as in Divjak and Gries (2006) or Finnish *think* verbs with either a singular first or third person subject/agent as in Arppe and Järvikivi (2007)), or to include all possible types of syntactic arguments of the immediate sentential context (e.g., Finnish *think* verbs, whether in finite or non-finite form, as in Arppe (2008)).

Using a corpus of written texts as our source of evidence, we can in the simplest form count raw frequencies of the near-synonyms studied among the contextual elements defined above and compare their proportions. If some proportional difference is statistically significant, we may infer that the contextual element in question differentiates the near-synonyms, so that relative frequency is associated with the preference of a lexeme. Resorting to experimental evidence of the off-line kind, we may instruct native speakers to rank the acceptability of the near-synonyms in

the sentential contexts of interest: if some rating difference is statistically significant, we may again infer that the associated contextual element is a differentiating one, but whether the element prefers or disprefers any of the lexemes has to be considered carefully, in comparison with the respective ratings of other contextual elements under study. As an alternative, on-line source of experimental evidence, we may ask native speakers to read word by word on a computer screen sentences containing contexts of interest and measure the durations allotted for each word: if some durational difference is statistically significant we may again infer that the associated contextual element differentiates the near-synonyms, probably so that longer duration entails the dispreference of a lexeme. Nevertheless, it is not a given that all of these resultant measures, (i.e., frequencies, proportions, ratings or durations), will yield convergent results because they are all manifestations of different linguistic activities – that is, writing, judgment and reading – and lead to different operationalisations of what is linguistically 'good' and what (perhaps) is not, (see, for example, Arppe and Järvikivi (2007) with respect to the comparison of corpus frequencies, forced-choice selections, and graded acceptability ratings). Even if the measures do converge, this does not necessarily mean that they reflect exactly the same underlying cognitive linguistic structures or processes.

The same problem pertains to the individual explanatory factors and the linguistic theoretical models that incorporate them, which need to be consistent from one type of evidence to the next. For instance, in our study of near-synonymy we may in the linguistic analysis of our research corpus follow, for example, the Functional Dependency Grammar formalism (Tapanainen and Järvinen, 1997) as in Arppe and Järvikivi (2007) and Arppe (2008). In order to retain the closest consistency in any subsequent experimentation, such as in acceptability ratings or reading time measurements, we need to apply the same theoretical formalism and follow the same definitions of the constituent categories in analysing the results so that we may properly compare the results from the different sources of evidence. Or, if, in the case of comparing data that was analysed earlier, we have to work with several theoretical formalisms, we need to develop a systematic way of mapping against one another the formalisms and the associated explanatory factors of interest. In particular, this is an issue when the original analyses of the different evidence types are undertaken by different researchers or groups and at different times.

The issue of cross-methodological consistency thus concerns, first, the question of how to operationalise a research question (and associated concepts) properly and consistently for each evidence type that one intends to use. To this end, one needs to understand sufficiently the fundamental characteristics and restrictions particular to each type of evidence as well as to be capable of using the appropriate state-of-the-art methods available for each – probably requiring the adaptation of the operationalisation in the process. Secondly, we need to ask how we can adequately compare and contrast the results gained with different methods.

While this task is difficult, some headway has been made. Recent results gained by multi-methodological linguistic research appear to be, if not fully convergent, at the very least reconcilable with one another. Studies that successfully combine different sources of data include, for instance, Gries (2002) and Rosenbach (2003) on the English genitive alternation, Gries *et al.* (2005) on the English *as*-predicative, Gries (2003), Bresnan (2007), and Bresnan *et al.* (2007) on the English dative alternation, Keller (2000) and Kempen and Harbusch (2005) on the middle field of German verbs, Jantunen (2004) and Vanhatalo (2005) on Finnish synonymous adjectives, and Arppe and Järvikivi (2007) on Finnish synonymous verbs. At the same time, different methods and types of evidence rarely yield exactly the same results. Nevertheless, even these differences help us gain a more accurate understanding of the linguistic phenomena studied, (i.e., what is actually happening behind the observed linguistic behaviour). Problems should not be an excuse for linguists to dig deeper into methodological monism: they are there to be solved.

The point that linguists should continue to look for converging evidence is supported by the fact that the replicability of results with different methods and types of evidence is actually the best measure of their robustness (Rosnow and Rosenthal, 1989; and Moran, 2003). Studies relying on converging evidence are in principle essentially more compelling than studies that merely increase the size of data sets of the same type.

Importantly, interdisciplinarity requires collaboration. Since mastering the methods of only one subfield can be challenging for a single researcher, practitioners of various subfields within linguistics should not only strive to co-operate with one another, but should also team up with researchers in the neighbouring disciplines. Much more collaborative research is necessary for us linguists to better understand the interrelationships between different methods and the evidence they produce.

2.2 Second thoughts on converging evidence

Dylan Glynn:
Converging evidence is an important goal, but there is a danger in comparing things that are not very well understood. There are fundamental questions, such as the relationship between frequency and salience, which we are yet to understand.

I believe the issue of converging evidence to be amongst the most important issues in contemporary linguistics. However, before we can speak of converging evidence, we first need to understand what is converging, and then we need to advance the state of the art in each domain sufficiently to offer viable results for comparison. Let us keep in mind that corpus linguistics is studying something very different from psycholinguistics. Also, both fields

are relatively new domains of research, especially for the model of language propounded by Cognitive Linguistics.

A first step is to agree upon the kinds of data that constitute empirical evidence. As I see it, we have essentially two kinds of empirical evidence at our disposal, which we can call 'found data' and 'elicited data'. These kinds of data are entirely different objects of study. With found data, we make generalisations based on relative frequency; with elicited data, we make generalisations based on relative salience. Gilquin (2008) more than adequately shows that the two do not necessarily correlate and it is my contention that we have little or no understanding of how results from these different types of data inform one another.

(*i*)   *Found data*. Found data commonly takes the form of text corpora. This kind of data directs research towards questions that can be answered through the observation of (relative) frequencies of occurrence. Such data can then yield generalisations about questions of natural language use – for example, how a form co-occurs with other forms, in what genres it occurs, and whether it displays sociolinguistic variation. Two basic methods can be used to treat such data:

    (*a*)   *Corpus analysis*. Qualitative analysis of the language existing in the corpus. Generalisations can be made inductively based on close inspection of linguistic forms.

    (*b*)   *Computational analysis*. Quantitative analysis of the language existing in the corpus. Patterns of language use can be proposed based on statistical (co-)occurrence tendencies of linguistic forms.

The main weakness of the first method is, of course, its inherent subjectivity. However, its strength lies in its ability to capture details that other methods miss. The second method is more objective, but is currently incapable of producing fine-grained results. An example of the first method is the corpus-based study of polysemy and near-synonymy (e.g., Divjak, 2006; and Glynn, 2009a). An example of the second method is the automatic disambiguation or word space modelling in computational linguistics (e.g., Gawronska and Erlendsson, 2005; Heylen *et al*. 2008b). The two kinds of method can be combined, an example of which would be collostructional analysis (Stefanowitsch and Gries, 2003).

(*ii*)   *Elicited data*. Secondly, we have elicited data that takes the form of speakers' recorded responses in either psycholinguistic experimentation or direct elicitation. Both methods face issues of representativeness.

    (*a*)   *Experimentation*. The use of quantitative methods of observation under controlled laboratory conditions to test hypotheses of language processing and use.

> (*b*)   *Interrogation*. Direct extraction of data by asking speakers or stimulating them to produce language.

The main strength of the first method is its objectivity. However, the results are always dependent on assumptions about the relationship between observable behaviour and mental activity. It is excellent for obtaining information on language processing and conceptual salience. The second approach is the mainstay of field linguistics, and it is indispensable for the production of reference grammars. It is difficult to obtain data through elicitation without the informants becoming aware of their own behaviour and hence making conscious linguistic choices. For some questions, it is imperative that speakers remain unaware of the real research question, which makes elicitation in these cases difficult or even unsuitable to apply.

Before we can speak of converging evidence, we must agree on what the evidence represents. The first step to doing this is to agree on the different kinds of data, then the different methods and what these methods and data tell us. I imagine that even my simplistic breakdown above would meet with a great deal of disagreement. In light of this, we must remain cautious about comparisons of results gleaned from found and elicited data. Let us first establish the different types of data, the methods for treating it, and then how data and methods relate to each other.

3. Corpus frequencies and psychological reality

3.1 Corpora are no shortcut to cognition

> *Gaëtanelle Gilquin*:
> A naive mapping of raw text frequency to cognitive salience is likely to fail. Corpora cannot stand in for experimental work.

Cognitive linguistics posits strong links between cognition and actual usage events, so that '[a]n event [. . .] becomes more and more deeply entrenched through continued repetition' (Langacker, 1987: 100). This assumption entails that corpora, which contain information about what is likely to be repeated or not in language, should make it possible to identify those items that have a special status in the mind. However, this assumption is mainly that – an assumption, and linguists have made relatively few efforts hitherto to test the cognitive reality of corpora.

One concept that may be used to test the relevance of text frequency to cognitive phenomena is prototypicality, according to which categories are organised around a maximally representative member (the prototype) and other members that are more or less representative of the category. The prototype was originally a psychological concept; it is supposed to be the 'most salient exemplar among the members of a category' (Radden, 1992: 519–20) and hence comes to mind first when people are asked about the

members belonging to this category (see Rosch's (1975) experiments on natural categories such as birds, vehicles and furniture). Using experiments similar to those performed by Rosch and colleagues, one can establish the most cognitively salient member of a linguistic category (arguably corresponding to its prototype) and compare it to the most frequent member as attested in a corpus. In some cases, the two measures do converge. The English preposition *through*, for example, turns out to be both most salient and most frequent in the construction [X moves through Y], where Y is a two-dimensional landmark (Gilquin and McMichael, 2008). Salience and frequency, however, do not always coincide (in fact, convergence seems to be the exception rather than the rule). Thus, while according to a sentence production experiment the verb *give* tends to be spontaneously associated with the meaning of 'handing over', its most frequent sense in the Switchboard corpus is the de-lexical sense (e.g., *give a smile*, *give a kiss*), a more collocational usage where the verb has little meaning of its own (Gilquin, 2008). Other studies have brought to light a similar discrepancy (e.g., Shortall, 2007; and Nordquist, 2009). In general, phraseological uses of words tend to be the most frequent usage patterns in corpora, whereas the most concrete senses tend to be highly salient.

The lack of convergence between salience and text frequency challenges the ability of corpora to serve as a shortcut to cognition. Two caveats should be introduced, however. The first one is that there may be methods of establishing frequency patterns that compare better to experimental studies than raw frequency. Ellis and Simpson-Vlach (2009), for instance, demonstrate that the mutual information (MI) of formulaic sequences determines processability more accurately than any other corpus-derived measure, including frequency. The second caveat is that it is possible, and in fact quite common, to study a concept like prototypicality on the basis of observable features such as historical origin, order of acquisition by children or, indeed, frequency of use in natural language – all of which may be determined more or less easily by means of corpus data. However, such criteria should be used with full awareness of the fact that they represent at best indirect effects of the cognitive phenomenon under study. At worst, they may not say anything about prototypicality as a cognitive phenomenon. In order to be able to make more reliable and psychologically relevant claims about cognition, one has to rely on a more direct manifestation of conceptualisation than language and go through the sometimes laborious process of experimentation.

3.2 Corpora and cognition revisited

Arne Zeschel:
Asking for psychological reality is not the same as asking for prototypicality, and corpus linguistics does not attempt to stand in for experimental work.

I agree that there are important unresolved questions when it comes to the interpretation of corpus findings as cues to psychological entrenchment properties (*i*). However, measures of 'prototypicality' may not provide the best yardstick to assess the relevance of corpora to cognitive linguistic research (*ii*). More generally, it is not obvious why experimental elicitation should be assumed to provide a necessarily more valid, direct, or otherwise privileged access to factors that are relevant for linguistic cognition than induction from naturally occurring productions (*iii*).

(*i*)   The usage-based hypothesis assumes that there is a connection between the usage frequency of linguistic structures and their degree of cognitive routinisation, or likelihood to be memorised/ stored (entrenchment). It also assumes that entrenchment is one of the three general factors that influence the choice of linguistic categorising structures during language processing (Langacker, 2000). Unless these assumptions are shown to be wrong (which, to my mind, the studies referenced above do not), data-driven identifications of structures that are likely to constitute entrenched units in a speech community on the basis of large representative corpora (arguably a problematic criterion) should be of interest to research on the cognitive instantiation of language. True, methodologically orientated work on the relationship between corpus results and language processing performance within cognitive linguistics (e.g., Gries *et al.*, 2005; and Wiechmann, 2008) has not established a global 'cognitive reality of corpora'. But since the impact of frequency on virtually all aspects of language processing and learning is hardly just an 'assumption' but, rather, a well documented fact (cf. Bybee and Hopper, 2001; and Ellis, 2002), corpus-derived findings are not necessarily in need of additional experimental corroboration in order to qualify as relevant for cognitive research. Indeed, if this is what one thinks, it is difficult to see why one would want to conduct a corpus study in the first place, rather than proceed straight to the experiment that is needed independently anyway.

(*ii*)   It has been pointed out in various places (e.g., Gilquin, 2006) that 'prototypicality' is multi-faceted and heavily ambiguous, or, in other words, that it is a seriously ill-defined term that is in need of substantial clarification in order to be more than just a conveniently vague 'catch-all' notion (Wierzbicka, 1985: 343). Furthermore, in contrast to the uncontroversial existence of prototype *effects* in categorisation, the psychological reality of prototypes-as-mental-representations is highly speculative and by no means uncontroversial in the categorisation literature (see Estes (1994) for in-depth discussion and comparison of different models). Hence, unless the concept is appropriately operationalised and, in the revised definition, shown to be

truly indispensable for the study of linguistic categorisation, I would hold that there is nothing inherently flawed about a method that does not 'say anything about prototypicality as a cognitive phenomenon', but, rather, uses much better-understood notions such as input frequency and statistically significant co-occurrence as cues to the psychological status of a given target structure.

(*iii*)   Gilquin (2008) teases apart two features that are often assumed to be indicative of prototypicality: salience (as measured by the relative accessibility of different senses of a polysemous word in a zero-context sentence production task) and frequency (as measured by sense frequency counts in two different corpora, Frown and Switchboard). Noting discrepancies between these measures in her data, the term 'prototypicality as a cognitive phenomenon' is then equated, in effect, with salience alone, which in turn prompts the conclusion that 'corpus analysis therefore cannot be used as a shortcut to cognition'. While the idea of critically evaluating alleged interdependencies between these and other presumed components of prototypicality is clearly the road to take if one is interested in this question, I believe that a study which compares production preferences in a particular experimental setting on the one hand, with textual frequency counts in two specific corpora on the other, does not necessarily warrant strong conclusions about inherent limitations of corpus data in the study of linguistic cognition as such. For one thing, in this particular study, the observation that the most frequent elicited sense of *give*, 'hand', is considerably less frequent in the Switchboard and Frown data does not necessarily show that it is in fact a merely secondary sense in discourse: arguably, a corpus of more natural kinds of spoken interaction (i.e., exchanges in which two or more interlocutors engage in face-to-face interaction within a shared speech situation) can be expected to contain higher proportions of 'hand'-usages (*Could you please give me the* X) than either displaced written communications (Frown) or likewise displaced telephone conversations between complete strangers (Switchboard). For instance, using the same semantic categories as Gilquin (2008) for annotation, a sample of the first fifty instances of *give* (in all relevant forms) in the spoken part of the BNC yields a proportion of 28 percent 'hand' senses, thus making it the most frequent sense in this spoken data sample. But suppose that the two measures do not, in fact, consistently yield the same results – which is indeed quite likely. Does this automatically undermine the validity of the corpus results as cues to psychological entrenchment patterns? Not necessarily. The fact that the most frequent corpus sense in the study (i.e., light verb *give* as in *give a shrug*) was not among the first that came to mind in the sentence production experiment may just as well

reflect a limitation of the experimental design rather than prove that frequency does not determine ease of activation: speakers tend to think of words as carriers of meaning, and word forms activate semantic frames; hence, when subjects are led to think about word meanings, it is perhaps not surprising that the most frequent responses do not involve semantically light to near-empty senses of the prime. To conclude, I do not believe that elicitation protocols provide an *a priori* more reliable probe into cognitive processes than other methods and that they can, therefore, serve as a gold standard for all other kinds of empirical research in cognitive linguistics.

4. Corpora and the study of alternations

4.1 Language is not a set of alternations

Dylan Glynn:
Our focus on alternations is the result of theoretical heritage from generative syntax and a matter of methodological convenience. Most linguistic decisions that speakers make are more complex than binary choices.

A large amount of the quantitative corpus research in cognitively orientated linguistics, especially research relying on confirmatory statistical techniques, has examined so-called alternations (Benor and Levy, 2006; Bresnan *et al.*, 2007; Diessel, 2008; Gries, 2002; Grondelaers *et al.*, 2008; Heylen and Speelman, 2003; Hilpert, 2008; Hinrichs and Szmrecsanyi, 2007; and Rosenbach, 2003; among many others). However, when a speaker chooses a concept, he or she chooses between a wide inventory of lexemes, each of which profiles different elements of that concept. Moreover, the speaker profiles that lexeme in combination with a wide range of grammatical forms, each also contributing to how the speaker wishes to depict the concept. In light of this very basic assumption in Cognitive Linguistics, alternations are as simplistic and reductionistic as the theories of language that were originally used to study them. Although the study of alternations has its place, it should be but the starting point of quantitative multi-factorial treatment of language. I argue that our preoccupation with this kind of research question is a result of two methodological errors, and that it therefore needs to be abandoned.

(*a*)   *Theoretical inheritance*. For their own theoretical purposes, the precursors of Cognitive Linguistics were concerned with what they called grammatical alternations. The fields of Cognitive Linguistics and Construction Grammar grew out of the battle to

disprove modularist, generative approaches to language. In order to demonstrate the shortcomings of these approaches, research focussed on phenomena that were studied in the generative literature. Although any attempt to refute the results of a given theory in order to support another is a good way of doing science, having achieved this, Cognitive Linguistics would now be better served by returning to its own model of language and focussing on its original research questions.

(*b*)   *Methodological convenience*. The second reason that we study alternations is because modern statistics is about falsifying a hypothesis – proving things wrong is the *raison d'être* of all confirmatory statistics. Although confirmatory statistics is not restricted to binary response variables, we tend to have relatively small data sets, and knowledge of multinomial regression methods remains beyond the knowledge of most corpus linguists. It is for these two practical reasons that working with binary response variables is preferred. However, statistical techniques should not be chosen out of convenience: one chooses a technique to match the research question in the most adequate way. Since language is not an inventory of linguistic alternations but a maze of inter-related linguistic choices and combinations, we need techniques that can model this complexity as well as determine the statistical significance and explanatory power of the model. Although Arppe (2008) and Glynn (2009b) have sought ways to achieve this, the statistical tools needed to model this degree of complexity remain beyond the current state of the art in linguistics.

4.2 Alternations are not everything, but they are useful

Gaëtanelle Gilquin:
There are good reasons to design research questions around binary response variables.

No one denies that language most of the time is more complex than a choice between two options. When speakers produce an utterance, this utterance has been selected from a theoretically infinite pool of other utterances, not just two alternative syntactic constructions. Further, no one denies that linguistic categories usually do not correspond to clear-cut dichotomies, but instead take the form of clines. This has become especially obvious with the advent of corpora. As Čermák (2002: 273) notes:

the historical scarcity of data (...) evoked the impression that language data is comfortably discrete and of an entity-like quality. What huge corpora show is rather different: most of the information is scalar,

obtainable in stepwise batches with hazy edges only, where the best help available is often statistics and fuzzy approaches and no longer black-and-white truths and clear-cut classification boxes. To put it differently, instead of insisting on getting straightforward answers of the *yes-no* type we have to elicit answers of the type *rather this than that*, or *more of this and less of that*.

Teubert (1996: v) says nothing different when he writes that '[o]ngoing research [in corpus linguistics] will replace many coarsely cut binary concepts by continuous scales'. This, however, does not mean that alternations and binary systems should – or indeed, could – be dispensed with. One reason for this is that alternations do exist: 'a construction, a word, or a morpheme is either present or absent, either has or doesn't have some formal property', as Givón (1992: 10) notes, referring to this as the 'discretizing' function of grammar. If a grammar encodes a difference, we should be paying attention to this difference. Even if a given alternation does not exhaust the logical space of what is possible in grammar, the study of alternations can nonetheless give us important clues about general organisational principles of language, as is evidenced by the very studies that Glynn mentions in his opening paragraph. For instance, we see the principle of end weight at work in both the dative alternation, where 'heavy' recipients favour the prepositional dative, and in binomials such as *salt and pepper*, where the lighter constituent usually comes first, rather than second. Another principle that is visible across different alternations is the so-called complexity principle (Rohdenburg, 2000), which predicts the use of a more analytic variant in situations that involve a greater processing load. This holds for the two English genitives as well as the two English comparatives. In the former case, abstract possessors are more difficult to process than concrete, animate possessors. Hence, the *of*-genitive is preferred with abstract possessors. In the latter case, a complement clause after the comparative adjective increases processing load and thus leads to increased usage of the periphrastic comparative. As a third example, the principle of sequential iconicity (i.e., the tendency to express first what happened first) is evident across different alternations. English temporal clauses with the subordinating conjunctions *before* and *after* exhibit a tendency to be ordered iconically relative to their main clauses. Similarly, sequential iconicity is at work in co-ordinated pairs such as *buy and hold*, *rise and shine* or *seek and destroy*.

    Finding and documenting such principles of linguistic organisation as they manifest themselves in frequency data is precisely what cognitive corpus linguistics should strive to do, and linguistic alternations provide a means to this end. While it is true that cognitive linguistics inherited the idea of alternations from generative linguistics, it would do itself a serious disservice by abandoning this kind of research. What we are doing with alternations now has little to nothing in common with the erstwhile goals of generative grammarians. As far as the argument from

methodological convenience is concerned, Glynn is right in demanding that research questions, not the availability of statistical methods, should lead the way. But given the rapid development of new corpus-linguistic methods that we have witnessed in recent years, it is probably too early to decry binary logistic regression as a block to further progress.

5. Experimental use of grammaticality judgments

5.1 Meta-linguistic tasks and their problems

> *Martin Hilpert*:
> Corpus linguistics has developed tools of remarkable precision and predictive power. Experimental validation of corpus-based results has become quite sophisticated as well, but we should strive to free it from meta-linguistic tasks.

Multi-factorial corpus-linguistic techniques allow corpus-based analyses of linguistic forms in terms of several explanatory variables and their interactions. Analyses of English data have addressed, among many other topics, the dative alternation (e.g., Gries, 2003; and Bresnan *et al*., 2007), the two genitives (e.g., Hinrichs and Szmrecsanyi, 2007), and the two comparatives (e.g., Mondorf, 2003; and Hilpert, 2008). The main appeal of these designs is their precision and predictive power. They generate results that are easily testable, and usually some such testing is carried out immediately by checking how well a model from Corpus A predicts the phenomenon under investigation in Corpus B. For instance, Bresnan *et al*. (2007) quite successfully use Switchboard corpus data to predict Wall Street corpus data.

Sometimes, the results reported in such studies sound a little too good to be true. If we learn that a given corpus model correctly classifies speakers' choices between two variants in 95 percent of all cases, this sounds like the problem has been almost completely solved and we now know what goes on in the minds of speakers when they choose one variant over the other. Given that corpora represent the language production of many speakers, and given that statistical models average over these aggregate data, it is probably natural to be sceptical and to question whether what holds true for a corpus will also allow us to predict the behaviour of a speaker 'in the flesh'. Whereas current statistical methods are able to control for the idiosyncrasies of individual speakers and writers in a corpus (similar to psycholinguistic by subjects analyses), many psycholinguists remain unconvinced that corpus data allow us to draw conclusions about cognition. In a recent issue of the *Zeitschrift für Sprachwissenschaft*, Matthias Schlesewsky (2009: 170) states that, in his view, corpus data are important but lack explanatory power:

Ich werde auch nichts über Korpusdaten schreiben, da ich sie zwar einerseits für wichtig halte, andererseits aber zumindest momentan überhaupt nicht sehe, welche explanatorische Kapazität über das Triviale hinaus sie besitzen
Neither will I be writing about corpus data, as on the one hand I consider them to be important, but on the other I fail to see at the moment what explanatory power beyond the trivial they may have.

[Translation by M.H.]

It appears that corpus linguists might have to do a better job at explaining why they work in the way they do. And there are less benign criticisms than the one offered by Schlesewsky. The reliance on aggregate data has prompted attempts to discredit the corpus-linguistic enterprise in its entirety. Newmeyer (2003) has argued prominently that '[g]rammar is grammar and usage is usage', thus denying that corpora have any explanatory value at all. The idea that corpus data are in need of some kind of cross-validation is thus not as outlandish as one might initially think – especially if one takes language to be a multi-faceted phenomenon, as Arppe argues above. Even corpus linguists who themselves do not perceive a need for converging evidence should have a professional interest in seeing this issue resolved.

In order to demonstrate the psychological reality of corpus-derived results, efforts have been made to validate them experimentally – typically using a dependent measure in the form of a judgment of grammaticality or acceptability. To illustrate, Gries (2003) asked subjects to rate typical, marginal, and atypical examples of both the ditransitive construction and the prepositional dative on a seven-point scale, finding that subjects' ratings confirmed the overall fit of his model. Rosenbach (2003) presented subjects with a forced choice between contextualised examples of the *s*-genitive and the *of*-genitive. Here, subjects' choices confirmed the hierarchy of three factors that had been found to be significant predictor variables in a prior corpus analysis. Bresnan (2007) modified this design and gave subjects a forced choice between the ditransitive construction and the prepositional dative and a simultaneous task to assess the relative likelihood of their choices. A linear regression of likelihood ratings showed the direction and relative magnitude of factors to be highly similar to the regression model from the corpus analysis. In my opinion, validations of this kind are examples of excellent research practice. However, I would like to argue that it is possible, and, indeed, preferable, to conduct validations that do not depend on judgments of acceptability.

The main issue I see is the question of how ecologically valid measures of grammaticality judgments are. Assessments of this kind always tap into meta-linguistic knowledge that is something quite different from the unconscious linguistic competence that linguistics tries to model. I would like corpus linguists to ask themselves the following question: what is the most natural measure of speaker behaviour that we could use to check corpus-generated results? There is, of course, a whole array of measures: reaction times, self-paced reading times, sentence completion, production

times, eye tracking, sorting tasks or memorisation tasks. Since corpora represent language production, I would like to argue that production-related measures such as relative degree of reduction or pronunciation length could be fruitfully used to replace acceptability judgment data: it is well-understood that linguistic units are produced in a reduced fashion (shorter overall, centralisation of vowels, omission of phonological elements) in contexts where they are highly predictable (Jurafsky *et al*., 2001). This does not only hold for collocations of specified elements, but generalises to co-occurrence patterns of lexical units with grammatical constructions (Gahl and Garnsey, 2004).

In a study that applies this insight to the validation of a corpus-based study, Tily *et al*. (2009) use Switchboard corpus recordings to test the regression model of Bresnan *et al*. (2007). The central question is whether speakers pronounce the elements *to* and *the* in a more reduced fashion when the examples they occur in are especially typical of the respective construction:

(1) *a.*   I gave the money *to* the Red Cross.
     b.   I gave Bob *the* car keys.

Tily *et al*. find a significant correlation of higher syntactic probability and reduced production with the prepositional dative, and a marginally significant correlation with the ditransitive. The overall approach of using production times in the place of acceptability judgments thus seems viable. It is important to point out that production measures of this kind can easily be transferred to other research questions. The two English comparatives may serve as a first example. Speakers of English exhibit a bias towards the periphrastic comparative in the context of a following *to*-infinitive clause (Mondorf, 2003). Hence, speakers should find example (2a) more 'natural' than example (2b) and produce a shorter token of *politer* in the first sentence.

(2) *a.*   He was *politer* to the dean.
     b.   It was *politer* to decline.

Of course, a following *to*-infinitive is just one variable among many others that govern the comparative alternation. A full production study would capture these variables across different sentences, determining whether production times are affected by these variables, and whether the effects that are observed converge with the predictions made by the corpus analysis.

Another possible application would be to measure production times in binomial expressions such as *brothers and sisters*, *sing and dance* or *cold and wet*, which display a preferred ordering that can be reversed (Benor and Levy, 2006). Frequencies in corpora can be used to predict just how reluctant speakers will be to utter the phrase *sisters and brothers*. Do we observe longer production times of the first element if it occurs in a dispreferred order? Are production times especially affected when the order militates against iconicity, end weight, or other constraints that we know of? In extending the

approach of Tily *et al*. (2009), one objective would be to design a method that could not only test the overall validity of a corpus model, but also the relative status of factors and interactions within that model. If we can reproduce the ranking of factors that is suggested by a corpus-based model, that would constitute a particularly strong kind of converging evidence.

Given the general viability of testing corpus models without recourse to acceptability ratings, I would like this route to become more widely adopted than it has been thus far. Having a protocol for testing corpus results with a canon of production measures will, in the long run, earn cognitively orientated investigations of corpora a much better reputation.

5.2 Meta-linguistic tasks are a form of language use

Antti Arppe:
If acceptability ratings provide us with consistent results that are reconcilable with corpus-based and other evidence, then we should use them.

As I argue above, I consider language to be a multi-faceted human phenomenon: language use encompasses a wide range of different linguistic activities. Thus, I am wary of classifying linguistic activities along a unidimensional scale, as presented for example in Gilquin and Gries (2009). The production of language in spontaneous conversation is arguably the most basic form of linguistic activity, but it is not the only form of language that should be taken seriously by corpus linguists. Linguistic judgments have been belittled as mere linguistic 'feelings' (Sampson, 2005: 17–18), or characterised as unnatural (Gilquin and Gries, 2009) – and, consequently, as having lesser value as linguistic evidence. To the contrary, I argue that making linguistic judgments is just as natural a linguistic activity as language production, albeit of a different quality.

Though we are not necessarily conscious of it, we constantly assess the linguistic output of our interlocutors, quickly noting if their particular form of language differs from ours. Distinguishing 'us' from 'them' on the basis of language usage is fundamental in establishing and defining the identity of a (linguistic or other) community. We are able to determine quite accurately whether a person we are talking to speaks a variety that differs from our own. Likewise, we immediately distinguish native from non-native speech, often determining the first language of a second-language speaker within seconds. By the same token, we clearly sense whether an utterance is correct or not in our own variety – or even in another variety. Although such naturally occurring linguistic judgments concern, first and foremost, spoken language, we can and do judge written texts in similar ways.

That the results of experiments based on grammaticality judgments are reconcilable and even convergent with corpus-based results is to my mind clear evidence that linguistic judgments are no less natural as a linguistic

activity than language production. Here, I take Tily *et al.*'s (2009) results to be an excellent case in point. Consequently, there is no pressing need to replace linguistic judgments as a form of evidence in linguistics. Rather, linguistic judgments should be used on an equal footing alongside evidence provided by other types of experimental methods, such as those building on language production.

6. Explanations in corpus linguistics

6.1 From descriptions to explanations...

Arne Zeschel:
Learning from positive evidence and statistical pre-emption can sufficiently explain grammaticality, but functionally oriented corpus linguistics should not only explain what the mental representations of speakers are, but also how the input for these representations came to be what they are today.

Construction-based approaches to grammar have introduced a new perspective on language: moving away from the traditional dichotomy of meaningful words and meaningless rules, language as a whole was recast as a system of uniform elements – constructions, first commonly defined as symbolic conventions that are in some respect non-predictable, later extended to any form–meaning pattern that is cognitively routinised. Along with this theoretical sea change came a turn to novel (or, at least within the tradition of Cognitive Linguistics, hitherto neglected) corpus-linguistic methodologies that permit a data-driven identification of these routinised patterns – the putative building blocks of speakers' linguistic knowledge. To be sure, 'characterizing the elements of speakers' linguistic knowledge' is a fair objective of linguistic research, and the combination of usage-based hypotheses with data-intensive methodologies clearly defines a promising arena for such research. Nevertheless, work in a paradigm that understands itself as 'cognitive–functional' in orientation should not content itself with such descriptions alone as part of an answer to the question *What is language made up of?*, but also seek to relate these descriptions to functionalist *explanations* as part of an answer to the question *Why is language the way it is?* Put differently, the ultimate goal of construction-based empirical research should not be to prove that the analysis of a given phenomenon requires the postulation of a particular construction in combination with a statistical identification of attendant entrenchment patterns. In addition, such studies should also seek to provide an account of *why* these patterns are the way they are, i.e., in how far they reflect the interplay of underlying cognitive, social, and interactional constraints on language use and language structure which caused these patterns to crystallise in the first place.

To date, however, these questions have mostly been addressed in isolation, rather than in combination. Depending on the interest of the analyst, the same phenomenon can be 'explained' in very different ways. Cognitive explanations will emphasise that the usage of a given form is governed by principles that ensure ease of production and processing. Sociolinguistic explanations will show that variation in usage can be predicted on the basis of extra-linguistic factors. Pragmatic explanations will reveal how the linguistic form is finely attuned to the particular needs of social interaction. Historical explanations will uncover the changes that produced the form in its current state and aim to reduce those changes in terms of general underlying principles of language change.

Whatever explanation one chooses: construction-based research should stop reiterating the same uncontroversial point about language being usage-based and entrenched patterns being likely to exhibit or develop all sorts of idiosyncrasies, and, instead, seek to offer integrated accounts that address issues of more general relevance that the investigated phenomenon can speak to. In short, characterising the mental representations that speakers use in linguistic categorisation remains an important goal of (Cognitive) linguistics – but there is more to Cognitive Linguistics than just this.

6.2 ... and back again

Martin Hilpert:
We can have our descriptions – and explanations, too.

It is probably an accurate observation that many corpus-based 'accounts' of grammatical constructions limit their argument to the following structure: first, we note the existence of a form that appears to have idiosyncratic properties; second, we use corpus data to analyse these properties, either gaining further empirical support or additional insights into the behaviour of the construction under investigation; and third, we repeat the well-rehearsed chorus that knowledge of language must be usage-based knowledge of constructions.

I agree with the point that Cognitive Linguistics has more to offer than this, but I do not think that cognitive linguists should stop producing papers that conform to the above schema. The cumulative study of constructions in a language goes beyond the mere establishment of an inventory. Rather, that process teaches us what functional categories are important in the language that we are looking at, and thereby constitutes a necessary preparation for any discussion of explanations – be they pragmatic, historical, social, cognitive or otherwise. To illustrate, corpus-based descriptions of polysemous constructions such as the ditransitive construction are necessary to suggest processes such as metaphor and metonymy as possible explanations of the polysemy we observe. If not for descriptions of constructions such as *let alone*, how would we understand the role of polarity in the grammar of English? Descriptions of idiosyncratic

constructions can be dismissed as mere grammatical butterfly-collecting, but such a dismissal overlooks the point that our understanding of grammar would be much poorer without them.

That said, descriptions and explanations are different things. Corpus-based accounts that show how functional pressures motivate form and meaning of a construction represent a different, and indeed indispensable, enterprise.

7. Outlook

The perspective taken in the arguments above is largely inward: what are the pressing questions that researchers within the discipline think about right now? The future of cognitive corpus linguistics partly depends on how these issues will be resolved, but it is perhaps just as important to consider its relation with the greater research community. Despite its widely recognised virtues, corpus linguistics is yet to be fully accepted as a fundamental method in Cognitive Linguistics. Reasons for this include:

(*i*) *A focus on form*. As a matter of tradition, corpus linguistics is primarily concerned with formal surface structures – collocations, patterns and the like. Although recent corpus-based research in Cognitive Linguistics has adopted corpus linguistic techniques to address issues in cognitive semantics, the image of a structure-driven, largely a-theoretical approach to language remains. If corpus methods are to gain any ground, it needs to be emphasised that these methods indeed speak to theoretical questions.

(*ii*) *A claim to objectivity*. Cognitive corpus linguistics is seen to take itself as more 'objective' than other methods of investigation. It is suggested that cognitive corpus linguistics somehow denies the paramount importance of subjective construal in language. This, of course, is not the case. We should make clear that we are interested in the very same issues that lie at the heart of Cognitive Linguistics writ large.

(*iii*) *A somewhat unbecoming arrogance*. The first two points result in a particularly unfortunate third one. Cognitive corpus linguistics should not dismiss intuition-based research, but concede that all hypotheses in fact begin with intuition. It is rare that a corpus study entirely nullifies the results of a good intuition-based study. Many corpus studies merely refine and improve the results of intuition-based studies.

In summary, corpus-based cognitive linguistics needs to do three simple things. First, a certain degree of humility could not hurt. Secondly, it should be shown how corpus linguistic techniques can answer the kind of research questions that cognitive linguists are traditionally interested in. Thirdly, the

real benefits of corpus methods should be stressed. These include the ability to produce testable results, the ability to handle complex phenomena, and the ability to connect with other areas of research to produce converging evidence. If these positive aspects are highlighted, the chances of corpus linguistics becoming the mainstream in Cognitive Linguistics should be increased.

References

Arppe, A. and J. Järvikivi. 2007. 'Every method counts – combining corpus-based and experimental evidence in the study of synonymy', Corpus Linguistics and Linguistic Theory 3 (2), pp. 131–59.

Arppe, A. 2008. Univariate, bivariate, and multivariate methods in corpus-based lexicography. A study of synonymy. Publications of the Department of General Linguistics, University of Helsinki, 44.

Benor, S.B. and R. Levy. 2006. 'The chicken or the egg? A probabilistic analysis of English binomials', Language 82, pp. 233–78.

Bresnan, J., A. Cueni, T. Nikitina and R.H. Baayen. 2007. 'Predicting the Dative Alternation' in G. Boume, I. Kraemer and J. Zwarts (eds) Cognitive Foundations of Interpretation, pp. 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, J. 2007. 'Is syntactic knowledge probabilistic? Experiments with the English dative alternation' in S. Featherston and W. Sternefeld (eds) Roots: Linguistics in Search of its Evidential Base, pp. 77–96. Berlin: Mouton de Gruyter.

Bybee, J. and P. Hopper (eds). 2001. Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins.

Čermák, F. 2002. 'Today's corpus linguistics. Some open questions', International Journal of Corpus Linguistics 7 (2), pp. 265–82.

Chafe, W. 1992. 'The importance of corpus linguistics to understanding the nature of language', in J. Svartvik (ed.) Directions in Corpus Linguistics, pp. 79–97. Berlin: Mouton de Gruyter.

Diessel, H. 2008. 'Iconicity of sequence: a corpus-based analysis of the positioning of temporal adverbial clauses in English', Cognitive Linguistics 19, pp. 457–82.

Divjak, D. 2006. 'Ways of intending: delineating and structuring near-synonyms', in St.Th. Gries and A. Stefanowitsch (eds) Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis, pp. 19–56. Berlin: Mouton de Gruyter.

Divjak, D.S. and St.Th. Gries. 2006. 'Ways of trying in Russian: clustering behavioral profiles', Linguistics and Linguistic Theory 2 (1), pp. 23–60.

Ellis, N.C. 2002. 'Frequency effects in language processing and acquisition', Studies in Second Language Acquisition 24, pp. 143–88.

Ellis, N.C. and R. Simpson-Vlach. 2009. 'Formulaic language in native speakers: triangulating psycholinguistics, corpus linguistics, and education', Corpus Linguistics and Linguistic Theory 5 (1), pp. 61–78.

Estes, W.K. 1994. Classification and Cognition. Oxford: Oxford University Press.

Gahl, S. and S.M. Garnsey. 2004. 'Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation', Language 80 (4), pp. 748–75.

Gawronska, B. and B. Erlendsson. 2005. 'Syntactic, semantic and referential patterns in biomedical texts: towards in-depth text comprehension for the purpose of bioinformatics' in B. Sharp (ed.) Natural Language Understanding and Cognitive Science. Proceedings of the second International Workshop on Natural Language Understanding and Cognitive Science (NLUCS 2005), pp. 68–77.

Gilquin, G. 2006. 'The place of prototypicality in corpus linguistics: causation in the hot seat', in St.Th. Gries and A. Stefanowitsch (eds) Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis, pp. 159–91. Berlin: Mouton de Gruyter.

Gilquin, G. 2008. 'What you think ain't what you get: highly polysemous verbs in mind and language' in J.-R. Lapaire, G. Desagulier and J.-B. Guignard (eds) Du fait grammatical au fait cognitif. From Gram to Mind: Grammar as Cognition (Volume 2), pp. 235–55. Pessac: Presses Universitaires de Bordeaux.

Gilquin, G. and St.Th. Gries. 2009. 'Corpora and experimental methods: a state-of-the-art review', Corpus Linguistics and Linguistic Theory 5 (1), pp. 1–26.

Gilquin, G. and A. McMichael. 2008. 'Measures of prototypicality: convergence or divergence? The case of *through*', Paper presented at the third International Conference of the German Cognitive Linguistics Association. September 2008. Leipzig, Germany.

Givón, T. 1992. 'The grammar of referential coherence as mental processing instructions', Linguistics 30, pp. 5–55.

Glynn, D. 2009a. 'Polysemy, syntax, and variation: a usage-based method for Cognitive Semantics' in V. Evans and S. Pourcel (eds) New Directions in Cognitive Linguistics, pp. 77–106. Amsterdam: John Benjamins.

Glynn, D. 2009b. 'Lexical fields, grammatical constructions, and synonymy: a study in usage-based Cognitive Semantics' in H.J. Schmid and S. Handl (eds) Cognitive Foundations of Linguistic Usage Patterns, pp. 89–118. Berlin: Mouton de Gruyter.

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

Gries, St.Th. 2002. 'Evidence in linguistics: three approaches to genitives in English' in R.M. Brend, W.J. Sullivan and A.R. Lommel (eds) LACUS Forum XXVIII: What Constitutes Evidence in Linguistics?, pp. 17–31. Fullerton: LACUS.

Gries, St.Th. 2003. 'Towards a corpus-based identification of prototypical instances of constructions', Annual Review of Cognitive Linguistics 1, pp. 1–27.

Gries, St.Th., B. Hampe and D. Schönefeld. 2005. 'Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions', Cognitive Linguistics 16 (4), pp. 635–76.

Gries, St.Th. and A. Stefanowitsch (eds). 2006. Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis. Berlin: Mouton de Gruyter.

Gries, St.Th. and M. Hilpert. 2009. 'From interdental to alveolar in the third person singular: a multifactorial, verb- and author-specific exploratory approach'. Paper presented at MMECL, University of Innsbruck.

Grondelaers, S., D. Geeraerts and D. Speelman. 2007. 'A case for a cognitive corpus linguistics' in M. Gonzalez-Marquez, I. Mittelberg, S. Coulson and M.J. Spivey (eds) Methods in Cognitive Linguistics, pp. 149–69. Amsterdam: John Benjamins.

Grondelaers, S., D. Speelman and D. Geeraerts. 2008. 'National variation in the use of er "there"' in G. Kristiansen and R. Dirven (eds) Cognitive Sociolinguistics. Language Variation, Cultural Models, Social Systems, pp. 153–203. Berlin: Mouton de Gruyter.

Hacking, I. 1996. 'The disunities of the sciences' in P. Galison and D.J. Stump (eds) The Disunity of Science: Boundaries, Contexts, and Power, pp. 37–74. Stanford: Stanford University Press.

Haspelmath, M. 2009. 'Welche Fragen können wir mit herkömmlichen Daten beantworten?', Zeitschrift für Sprachwissenschaft 28 (1), pp. 157–62.

Heylen, K. and D. Speelman. 2003. 'A corpus-based analysis of word order variation: the order of verb arguments in the German middle field' in D. Archer, P. Rayson, A. Wilson and T. McEnery (eds) Proceedings of the Corpus Linguistics 2003 conference, pp. 320–9. Lancaster University: UCREL.

Heylen, K., J. Tummers and D. Geeraerts. 2008. 'Methodological issues in corpus-based cognitive linguistics' in G. Kristiansen and R. Dirven (eds) Cognitive Sociolinguistics. Language Variation, Cultural Models, Social Systems, pp. 91–128. Berlin: Mouton de Gruyter.

Heylen, K., Y. Peirsman and D. Geeraerts. 2008. 'Modelling Word Similarity. An Evaluation of Automatic Synonymy Extraction Algorithms' Proceedings of the Language Resources and Evaluation Conference (LREC 2008), pp. 3243–9.

Hilpert, M. 2008. 'The English comparative: language structure and language use', English Language and Linguistics 12 (3), pp. 395–417.

Hinrichs, L. and B. Szmrecsanyi. 2007. 'Recent changes in the function and frequency of standard English genitive constructions: a multivariate analysis of tagged corpora', English Language and Linguistics 11 (3), pp. 437–74.

Jantunen, J.H. 2004. Synonymia ja käännössuomi: korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin. Ph.D. thesis, University of Joensuu.

Jurafsky, D., A. Bell, M. Gregory and W.D. Raymond. 2001. 'Probabilistic relations between words: evidence from reduction in lexical production' in J. Bybee and P. Hopper (eds) Frequency and the Emergence of Linguistic Structure, pp. 229–54. Amsterdam: John Benjamins.

Keller, F. 2000. Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality. Unpublished Ph.D. thesis, University of Edinburgh.

Kempen, G. and K. Harbusch. 2005. 'Grammaticality ratings and corpus frequencies' in S. Kepser and M. Reis (eds) Linguistic Evidence: Empirical, Theoretical and Computational Perspectives, pp. 329–49. Berlin: Mouton de Gruyter.

Langacker, R.W. 1987. Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites. Stanford: Stanford University Press.

Langacker, R.W. 2000. 'A dynamic usage-based model' in M. Barlow and S. Kemmer (eds) Usage-based Models of Language, pp. 1–63. Stanford: CSLI.

Mondorf, B. 2003. 'Support for More-Support' in G. Rohdenburg and B. Mondorf (eds) Determinants of Grammatical Variation in English, pp. 251–304. Berlin: Mouton de Gruyter.

Moran, M.D. 2003. 'Arguments for rejecting sequential Bonferroni in ecological studies', OIKOS 100 (2), pp. 403–40.

Newmeyer, F. 2003. 'Grammar is grammar and usage is usage', Language 79, pp. 682–707.

Nordquist, D. 2009. 'Investigating elicited data from a usage-based perspective', Corpus Linguistics and Linguistic Theory 5 (1), pp. 105–30.

Radden, G. 1992. 'The cognitive approach to natural language' in M. Pütz (ed.) Thirty Years of Linguistic Evolution. Studies in Honour of

René Dirven on the Occasion of his Sixtieth Birthday, pp. 513–41. Amsterdam: John Benjamins.

Rohdenburg, G. 2000. 'The complexity principle as a factor determining grammatical variation and change in English' in I. Plag and K.P. Schneider (eds) Language Use, Language Acquisition and Language History: (Mostly) Empirical Studies in Honour of Rüdiger Zimmermann, pp. 25–44. Trier: Wissenschaftlicher Verlag.

Rosch, E. 1975. 'Cognitive representations of semantic categories', Journal of Experimental Psychology 104 (3), pp. 192–233.

Rosenbach, A. 2003. 'Aspects of iconicity and economy in the choice between the s-genitive and the of-genitive in English' in G. Rohdenburg and B. Mondorf (eds) Determinants of Grammatical Variation in English, pp. 379–411. Berlin: Mouton de Gruyter.

Rosnow, R. and R.A. Rosenthal. 1989. 'Statistical procedures and the justification of knowledge in psychological science', American Psychologist 44 (10), pp. 1276–84.

Sampson, G. 2005. 'Quantifying the shift towards empirical methods', International Journal of Corpus Linguistics 10 (1), pp. 15–36.

Sampson, G. 2007. 'Grammar without grammaticality', Corpus Linguistics and Linguistic Theory 3 (1), pp. 1–32.

Shortall, T. 2007. Cognition, Corpus, Curriculum. Unpublished Ph.D. thesis, University of Birmingham, Birmingham.

Schlesewsky, M. 2009. 'Linguistische Daten aus experimentellen Umgebungen: Eine multiexperimentelle und multimodale Perspektive', Zeitschrift für Sprachwissenschaft 28 (1), pp. 169–78.

Schütze, C.T. 1996. The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology. Chicago: University of Chicago Press.

Stefanowitsch, A. and S.Th. Gries. 2003. 'Collostructions: investigating the interaction between words and constructions', International Journal of Corpus Linguistics 8 (2), pp. 209–43.

Stefanowitsch, A. and S.Th. Gries (eds). 2006. Corpus-based Approaches to Metaphor and Metonymy. Berlin: Mouton de Gruyter.

Tapanainen, P. and T. Järvinen. 1997. A non-projective dependency parser. In Proceedings of the fifth Conference on Applied Natural Language Processing, Washington, D.C., pp. 64–71. April 1997. Association of Computational Linguistics.

Teubert, W. 1996. 'Editorial', International Journal of Corpus Linguistics 1 (1), pp. iii–x.

Tily, H., S. Gahl, I. Arnon, N. Snider, A. Kothari and J. Bresnan. 2009. 'Pronunciation reflects syntactic probabilities: evidence from spontaneous speech', Language and Cognition 1 (2), pp. 147–65.

Tummers, J., K. Heylen and D. Geeraerts. 2005. 'Usage-based approaches in Cognitive Linguistics: a technical state of the art', Corpus Linguistics and Linguistic Theory 1 (2), pp. 225–61.

Vanhatalo, U. 2005. Kyselytestit synonymian selvittämisessä – Sanastotietoutta kielenpuhujilta sähköiseen sanakirjaan. Ph.D. thesis (published), University of Helsinki.

Wiechmann, D. 2008. 'On the computation of collostruction strength: Testing measures of association as expressions of lexical bias', Corpus Linguistics and Linguistic Theory 4 (2), pp. 253–90.

Wierzbicka, A. 1985. Lexicography and Conceptual Analysis. Ann Arbor: Karoma.