Corpora and experimental methods: A state-of-the-art review

GAËTANELLE GILQUIN and STEFAN TH. GRIES

Abstract

This paper offers a state-of-the-art review of the combination of corpora and experimental methods. Using a sample of recent studies, it shows (i) that psycholinguists regularly exploit the benefits of combining corpus and experimental data, whereas corpus linguists do so much more rarely, and (ii) that psycholinguists and corpus linguists use corpora in different ways in terms of the dichotomy of exploratory/descriptive vs. hypothesis-testing as well as the corpus-linguistic methods that are used. Possible reasons for this are suggested and arguments are presented for why (and how) corpus linguists should look more into the possibilities of complementing their corpus studies with experimental data.

Keywords: *Corpus linguistics; psycholinguistics; corpora; experiments; frequency; concordance; collocation; judgments; intuition.*

> A linguist half-sees that it would be convenient for him if some particular, fairly unusual sequence of words were grammatical, perhaps because it enables him to make some part of his grammar of English especially elegant, or because it constitutes a counter-example to some well-entrenched theory of universals and thus leads to fame for him as the David who overturns the theory; he mulls the word-sequence over in his mind for a while and pretty soon, lo and behold! he perceives (quite sincerely) a clear intuitive conviction that the string is indeed grammatical (in 'his dialect'). Sampson (1980: 152)

1. Data in linguistics: Some background

As in any proper scientific discipline, data are central to linguistics. Ever since linguistics materialized as a discipline in its own right linguists have been concerned with data of various sorts. To date, however, there is surprisingly little agreement on what exactly qualifies as data and how they are

Corpus Linguistics and Linguistic Theory 5–1 (2009), 1–26
DOI 10.1515/CLLT.2009.001
1613-7027/09/0005-1

to be obtained, analyzed, evaluated, and interpreted. European comparative linguists in the first half of the 20th century and American structuralists alike were data-oriented linguists who collected their data in their natural habitats, so to speak. For instance, Bloomfield adopted a data-driven approach, which Harris (1993: 27) describes as follows: "The approach [. . .] began with a large collection of recorded utterances from some language, a corpus. The corpus was subjected to a clear, stepwise, bottom-up strategy of analysis". This quote is also interesting because the word *corpus* is used to refer to data collected and elicited in fieldwork situations in communities where note-taking (or, nowadays, recording) and prompting native speakers can be rather invasive and remote from natural communicative settings. This contrasts with the modern ideal of a corpus to comprise as much natural language as possible.

As is well known, a fundamental change of perspective was introduced with the rise of formal syntax, or more specifically, transformational-generative grammar, in the 1950s and 1960s. This went hand in hand with abandoning collected data of the above kind and introducing very informally collected linguistic acceptability judgments (largely by the analyst him/herself) as the primary source of data. Paradoxically, the lack of sophistication adopted in gathering the judgments was considered convenient and a virtue:

> The gathering of the data is informal; there has been very little use of experimental approaches (outside of phonetics) or of complex techniques of data collection and data analysis of a sort that can be easily devised, and that are widely used in the behavioral sciences. The arguments in favor of this seem to me quite compelling; basically, they turn on the realization that for the theoretical problems that seem most critical today, it is not at all difficult to obtain a mass of crucial data without the use of such techniques. Consequently, linguistic work, at what I believe to be its best, lacks many of the features of behavioral sciences. (Chomsky 1969: 56, quoted from Schütze 1996: 5)

Given such a position, it is hardly surprising that judgment data quickly enjoyed unparalleled primacy in theoretical syntax and semantics. Data that conflicted with a linguist's own judgments were simply discarded (see Wasow and Arnold [2005: 1485] for a striking example).[1]

Thankfully, these areas have been undergoing a change of perspective recently, and the vast array of problems of this methodological orientation is now more openly discussed. Schütze (1996) was maybe the first study to address a multitude of factors that influence acceptability judgments including subject-related factors (e.g., field dependence, handedness, linguistic training, educational level) as well as task-related factors (e.g., order of presentation, frequency of exposure, self-awareness, imposed speed of judgment, frequency of the stimulus in the language, lexical content, truth values); more work has been done in this area since then. While the results are too numerous and diverse to be recapitulated here, it is fair to say that they show that accept-

ability judgments are data; at the same time, however, they tend to be noisy, volatile, less objective, and less generalizable than previously assumed.

On the other hand, abandoning judgment data altogether would mean throwing out the baby with the bathwater. While informally collected judgment data exhibit a variety of problems, judgment data collected with all the necessary precautions can be a valuable tool. Unfortunately, many linguists, rather than obtaining judgment data the careful way, are still preoccupied with general discussions of the pros and cons of introspective judgment data or, worse even, defenses of informally collected data. Especially with regard to the latter, several arguments are repeatedly provided (a good recent overview is Borsley 2005). Most of these arguments can easily be refuted, and we will briefly address three of them.

First, there is the 'but-we-admitted-it-all-along' counterargument. In defense of introspective judgments, Borsley (2005: 1476) adduces the following much-quoted passage by Chomsky:

> It is not that these introspective judgements are sacrosanct and beyond any conceivable doubt. On the contrary, their correctness can be challenged in various ways, some quite indirect. Consistency among speakers of similar backgrounds, and consistency for a particular speaker on different occasions is relevant information. The possibility of constructing a systematic and general theory to account for these observations is also a factor to be considered in evaluating the probable correctness of particular observations. (1964: 79–80)

If one looks carefully through Chomsky's writings over the last 40 years, one can probably find passages supporting, or at least not ruling out, virtually every position on data and methodology in linguistics. Accordingly, the sole purpose of the passage above seems to be to disarm critics of introspective judgment data with the correct (!) observation that such data are problematic; it only pays lip service to diversity and consistency. In fact, Chomsky himself never sought consistency in judgment data that were obtained using the careful experimental designs common in psycholinguistic studies (cf. Wasow and Arnold 2005: 1483–1484). Labov (1975: 100–101) summarizes Chomsky's approach as follows:

> When Chomsky encounters disagreement on intuitions he frequently notes the fact: for example, in discussing *our election of John* (acceptable) vs. *our election of John president* (unacceptable), he notes 'Reactions to these sentences vary slightly: [these] represent my judgment' (1973). He then continues, 'Given such data. . . . ' The data which Chomsky refers to is not the fact that reactions vary, but rather his own judgments, and he proceeds to argue on the basis of these alone.

Second, there is the 'but-you-do-it-too' counterargument. Borsley (2005: 1477) counters Sampson (1975) and Stubbs (2001), arguing that they can only know from intuition that *All apples are round* implies *This apple is round* and that the events referred to by *get* passives are often unpleasant. Nobody

would deny that most linguists rely on their intuition at certain stages of analysis. The crucial difference is that the judgment that *All apples are round* implies *This apple is round* is not likely to depend on the amount of exposure to both sentences, age, sex, general intelligence, profession, etc. While the empirically accurate way would be to run an experiment to test this, if one was forced to decide whether *John didn't leave until midnight but Bill did* is acceptable or not (Grinder and Postal's [1971] examples) or whether *All apples are round* implies *This apple is round* or not, it is obvious which of the two questions would elicit more diverse responses. The fact that especially the generative literature is replete with various numbers of asterisks, question marks, and combinations thereof underscores that point (cf. Wasow and Arnold [2005: 1482] for a similar argument). When the only source of data is a single linguist's intuition, that procedure is reminiscent of "the procedure of attempting to establish a case on the basis of a set of data the size of a small workbook problem (though with theoretical biases of more generous proportions)", in Pullum's (1978: 400) words.

Finally, there is the 'judgments-are-not-*invented*-data' counterargument. In response to Stubbs (1996), Borsley (2005: 1477, our italics) states that "[t]he sentences that linguists investigate may well be invented, *but the speaker's judgements are not invented* and it is these that are the data with which theoretical linguists work". However, the italicized part is most problematic – again, Wasow and Arnold (2005) and Pullum (2007: 37–38) are most instructive in this regard. The former provide examples where judgments about a particular idiom are simply asserted and clearly contradicted by corpus data; the latter discusses a stunning case where even within one and the same reference syntactically perfectly identical expressions get so wildly differing judgments that they are hard to explain without recourse to the word *invented*.

These and other considerations have led to a different empirical culture in many areas of linguistics, one that embraces different kinds of evidence rather than relying on intuitions alone. In fact, areas of linguistics other than theoretical syntax and semantics have been methodologically more diverse for quite some time already. For example, empirically more diverse and robust approaches are established in phonetics, sociolinguistics, psycholinguistics, and corpus linguistics. The latter two approaches have not only rapidly grown into disciplines in their own right over the last two or three decades (with discipline-specific questions, terminologies, and methodologies as well as approaches to data), but there are now also many studies that combine experimental and corpus data. The papers in this special issue are among the most recent examples of this trend. Let us continue to explore the full range of data that linguists can work with before we survey each of these papers in turn.

> Looking back at the syntax published a couple of decades ago makes it rather clear that much of it is going to have to be redone from the ground up just to reach minimal levels of empirical accuracy. Faced with data flaws of these proportions, biology journals issue retractions, and researchers are disciplined or dismissed. Pullum (2007: 36)

2. Different kinds of linguistic data

Apart from the above-mentioned points of critique, the prevalence of introspective judgment data is even more surprising once we look at the vast range of data that linguists theoretically have at their disposal. Table 1 provides an overview (at a high level of granularity, of course), roughly in descending order of naturalness of data production/collection (cf. Tummers et al. 2005 for a different characterization).

Table 1. *Kinds of linguistic data (sorted according to naturalness of production/collection)*

#	Data source
1	corpora with written texts (e.g., newspapers, webblogs)
2	example collections
3	corpora of recorded spoken language in societies/communities where note-taking/recording etc is not particularly spectacular/invasive
4	corpora with recorded spoken language from fieldwork in societies/communities where note-taking/recording etc is spectacular/invasive
5	data from interviews (e.g., sociolinguistic interviews)
6	experimentation requiring subjects to do something with language they usually do anyway, e.g., – tsentence production as in answering questions in studies on priming – tpicture description in studies on information structure
7	elicited data from fieldwork (e.g., responses to "how do you say X in your language?")
8	experimentation requiring subjects to do something with language they usually do not do, on units they usually interact with, e.g., – sentence sorting – measurements of reaction times in lexical decision tasks – word associations
9	experimentation requiring subjects to do something with language they usually do not do on units they usually interact with, involving typical linguistic output, e.g., – measurements of event-related potentials evoked by viewing pictures – eye-movements during reading idioms – acceptability/grammaticality judgments on units they usually do not interact with, involving the production of linguistic output, e.g., – phoneme monitoring – gating – ultrasound tongue-position videos

As is obvious, Table 1 makes reference to corpora or corpus-like data several times but on different levels of naturalness. It is therefore necessary to briefly explain what exactly we mean by *corpora*. Our approach is similar to Gries's (2006: 4–5) radial-category approach to defining *corpus linguistics*: there are several criteria that, if met, define a prototypical corpus, but the criteria are neither all necessary nor jointly sufficient. For us, a corpus is a collection of texts that

- is *machine-readable*;
- is *representative* with regard to a particular variety/register/genre, meaning that the corpus contains data for each part of the variety/register/genre the corpus is supposed to represent;
- is *balanced* with regard to a particular variety/register/genre, meaning that the corpus parts' sizes are proportional to the parts of the variety/register/genre the corpus is supposed to represent (given the absence of reliable estimates of how much of a target language consists of any one particular variety/register/genre, balancedness is a theoretical ideal);
- has been produced in a *natural communicative setting*.

This implies that, as reflected in Table 1, there is actually no strict corpora-experiments dichotomy. Rather, just as linguistic data in general form a continuum of naturalness of production/collection, so do corpora: they vary along the above dimensions, which results in a continuum ranging from prototypical corpora via less typical corpora to corpora whose compilation is distinctly experimental in nature.

An example of a prototypical corpus is the British National Corpus, which is machine-readable, has been compiled with an eye to including very many different registers and genres of the target language, and contains spoken and written data that have been produced in natural communicative settings. Corpora such as Brown and LOB are similar in this regard. While they are less comprehensive in terms of the registers they include – neither contains spoken language – they are both based on an elaborate sampling scheme that aims at ensuring representativity for their target varieties. The difference is therefore one in scope, but not in corpus quality or typicality as defined here.

Corpora that are slightly less prototypical along the above dimensions would be:

- the Switchboard corpus (Godfrey and Holliman 1997), which contains telephone conversations between strangers on assigned topics; while talking on the phone is a normal aspect of using language, talking to strangers about assigned topics is not;
- the International Corpus of Learner English (Granger et al. 2002), which contains timed and untimed essays written by foreign language learners of

English on assigned topics; while writing about a topic is a normal aspect of using language, writing on an assigned topic under time pressure is not.

While these corpora are not prototypical according to our definition, they are still close to the prototype. Thus, even if talking to strangers about assigned topics on the phone is not part of what we normally do, this is still a fairly regular way of using language: we do talk to strangers on the phone often; often, the topic is limited in a similar way; and the communicative behavior is still 'normal' in the sense that the participants try to communicate as if in a normal communication setting. And while writing on an assigned topic under time pressure is not per se a normal aspect of using language, it is quite natural in the context in which the ICLE data were collected, i.e. the context of the EFL classroom.

In some sense, corpora consisting of newspaper texts – journalese – are even more remote from what we defined as a prototypical corpus. While many linguists use such corpora (certainly at least partly for the sake of size and convenience), newspaper articles are a very particular register: they are created much more deliberately and consciously than many other texts, they often come with linguistically arbitrary restrictions regarding, say, word or character lengths, they are often not written by a single person, they may be heavily edited by editors and type setters for reasons that again may or may not be linguistically motivated, etc.

Finally, experimental corpora differ even more markedly from the typical naturalness of the communicative setting. Connine et al. (1984) collected sentences from subjects in an experimental setting where the subjects were prompted to write down sentences using words from a list and on a given topic or setting. Similarly, Garnsey et al. (1997) collected sentences subjects wrote when prompted to complete a sentence fragment. Obviously, both kinds of situations are rather remote from anything that might be called natural communicative settings. For yet another kind of more extreme deviation from prototypical corpora, consider the DCIEM Map Task Corpus (Bard et al. 1996). This corpus consists of task-oriented, but unscripted dialogs in which one interlocutor describes a route on a map to the other after both interlocutors were subjected to 60 hours of sleep deprivation and to one of three drug treatments. The communicative situation is maybe more regular – giving directions is not an untypical activity even though the maps here were rigged – but other aspects of the situation are obviously more peculiar to this particular study. While some corpus linguists might not even want to call these data corpora anymore, we prefer to be less prescriptive and/or judgmental and include them as corpora, if only less prototypical ones.

In addition to what we consider to be *corpora*, we should also say something about what we consider to be an experiment. There are a few examples in which *experiment* is used to refer to corpus queries. We prefer a different

terminology that is more in sync with the time-honored distinction between experimental and observational studies. We consider experimental studies to be studies where independent variables are manipulated systematically to determine what effect, if any, they have on a (set of) dependent variable(s), and which, in the humanities and social sciences at least, come with often considerable variation between subjects, stimuli, stimulus presentations, and, occasionally at least, with a mismatch between high internal validity but lower external validity because of the occasionally artificial settings experiments tend to require – at the same time, these settings of course minimize noise.

On the other hand, corpus studies are observational studies because variable levels are not assigned to cases before the study, but collected/recorded after the fact on the basis of the observed data. In addition, the data do not undergo variation at the time the researcher queries the corpus: if a query is repeated, it will return the same results over and over again, whereas every different subject or repeated tests on an already tested subject will most likely yield at least slightly different results. Finally, in many cases external validity is extremely high because the data were collected in natural communicative settings, which in turn increases the amount of noise in the data. Given all these differences, using the term *experiment* for corpus queries is somewhat misleading.

3. Corpora and psycholinguistic experiments

With regard to corpus linguistics, opinions differ as to whether it is a theory or 'just' a set of methods or tools. For our present purposes, it is not really necessary to commit to one of the two positions – we favor the latter and refer to Taylor (2008) for a recent brief overview – but it is worth pointing out that especially proponents of the former position sometimes appear to overrate corpus data in a similar, though usually less extreme, way that most theoretical syntacticians overvalued judgment data. Our position on this is that corpora are a supreme tool for linguistic data analysis with several well-known advantages, which follow from our above characterization:

- the data are from natural contexts; thus, they make it possible to study register/genre questions that are difficult to study experimentally and come with a higher degree of external validity than some experimental designs;
- a larger range of data can be investigated than many experimental designs allow for; for example, if 10,000 hits of a particular argument structure construction are studied, the number of verbs in that construction is probably higher than anything that can be studied in an experiment; if thousands of potential cases of syntactic priming are studied, the numbers of different

distances between primes and targets is larger than what can be included in an experiment;
– the diversity or noisiness of the data can nowadays be handled much better with multifactorial statistics (in particular exploratory approaches, but also generalized mixed effects or multilevel models).

However, like any other method in isolation, corpora are not perfect. This recognition has resulted in a flurry of recent studies that combine methodologies, preferably corpus and experimental data. The following are a few well-known advantages of experimental approaches:

– they allow the study of phenomena that are too infrequent in corpora, and given the Zipfian distribution so characteristic of linguistic data, these are more than one might expect;
– they make it possible to systematically control for confounding or moderator variables;
– depending on the nature of the experiment, they permit the study of online processes.

Because the advantages and disadvantages of corpora and experiments are largely complementary, using the two methodologies in conjunction with each other often makes it possible to (i) solve problems that would be encountered if one employed one type of data only and (ii) approach phenomena from a multiplicity of perspectives, as will be briefly exemplified below. In what follows, we would like to examine the role and use of corpora in a sample of recent studies, and show how they can be fruitfully combined with experimental methodologies. The first sample contains 'corpus-only' studies, whereas the second sample includes studies combining corpus data with experiments.

3.1. What corpus linguists do with corpora

In order to study the way corpus linguists use corpora, we collected the most recent issues (2005–) of three journals representative of the discipline, namely the *International Journal of Corpus Linguistics* (issues 10.1 to 13.2), the *ICAME Journal* (issues 29 to 32), and the present journal, *Corpus Linguistics and Linguistic Theory* (issues 1.1 to 3.2).[2] We included only those papers that actually analyzed corpus data (to the exclusion of experimental data),[3] and thus disregarded articles having to do with issues such as corpus compilation (e.g., Santini 2006) or theoretical position (e.g., Teubert 2005), which is of course not to downplay the importance of such topics in the field of corpus linguistics. The remaining papers, 81 in total, were categorized according to several criteria:

- approach: exploratory/descriptive vs. hypothesis-testing;
- topic of investigation: phonological vs. morphological vs. syntactic vs. semantic/lexical vs. pragmatic;
- perspective: corpus-linguistic vs. psycholinguistic vs. computational-linguistic;
- type of corpus 1: spoken vs. written;
- type of corpus 2: general vs. specific;
- corpus methodology: frequency vs. collocation/co-occurrence/association vs. concordance (or any combination of the three).[4]

It turns out from the examination of the 'corpus-only' sample that corpus linguists predominantly approach corpus data in an exploratory fashion, i.e., without rigorously formulated hypotheses made explicit in the paper (72% of the studies). In other words, they seem to favor a corpus-driven approach, which 'lets the data speak for themselves'. As far as the topic of investigation is concerned, the majority of the papers deal with lexis (60%) – either on its own or in combination with another topic – especially phraseological issues (collocations, idioms, semantic prosody, etc). 41% of the papers analyze a syntactic phenomenon. In comparison, morphology, pragmatics, and phonology represent a small proportion of the topics investigated (about 7% each). Studies exploiting a written corpus are more frequent than studies exploiting a spoken one (77% vs. 58%), and specific corpora (especially newspapers) are slightly more common than general corpora (59% vs. 52%). Unfortunately, this is probably more a reflection of the availability of corpora or the ease with which data can be collected than a theoretically motivated decision or even necessity. Finally, a look at the methodology used reveals an almost equal preference for the use of frequency and concordance (60% for the former and 58% for the latter), followed by the use of collocation/co-occurrence/association (32%). As the percentages already indicate, these methodologies are regularly combined with each other. Thus, the combination of frequency and concordance accounts for over 17% of the papers from our sample. Particularly striking is the combination of all three main methods (concordance, frequency, and collocation), which represents 12% of the studies and may be illustrated by Thompson and Sealey's (2007) study of the features of children's literature, which investigates the most frequent words and sequences of words/parts-of-speech in a corpus of fiction written for children, but also examines the concordance lines in which the frequent items occur.

3.2. What psycholinguists do with corpora

We were then interested to find out how corpora are used in combination with experimental methods. The sample studies on which we base our observations were retrieved by means of two bibliographical databases, namely

Scopus (http://www.info.scopus.com) and *Linguistics and Language Behavior Abstracts* (http://www.csa.com/factsheets/llba-set-c.php). The keywords we used were i 'corpus' AND 'experiment', and ii 'corpus' AND 'elicitation'. We restricted our selection to those papers that were available electronically,[5] whose abstracts seemed to point to an actual combined use of corpus and experimental data, and which were written in one of the languages we understand. Not all of the 104 papers thus retrieved made their way into our final sample. First, in a number of cases the terms *corpus* or *experiment* were used in the paper in a way that did not correspond to our definitions. For example, Kliegl's (2007) *corpus* actually refers to an eye-fixation corpus, and Martin et al.'s (1996) to an example collection (of semantic errors produced by normal and aphasic speakers on a picture naming test). Second, some papers adopted a purely theoretical perspective (e.g., Borsley 2005), only mentioned corpus data on the side (e.g., Greenbaum 1976), or were review/overview papers (e.g., Branigan et al. 1995 or Lippmann 1997). Such papers, nineteen in total, were disregarded. The remaining 85 papers constitute our 'corpus and experimental' sample and were categorized according to the same criteria as the ones discussed in Section 3.1 plus the following two:

– experimental methodology, e.g., sentence production, eye-tracking, lexical decision;
– role of corpus: validator (i.e., the corpus serves as a validator of the experiment) vs. validatee (i.e., the corpus is validated by the experiment) vs. equal (i.e., corpus and experimental data are used on an equal footing) vs. stimulus composition (i.e., the corpus serves as a database from which fitting examples are culled).

In this section, we deal with the studies from this sample that adopt a psycholinguistic orientation. The first thing to notice is that psycholinguistically oriented studies represent the large majority of our 'corpus and experimental' sample. They account for 78% of the sample, as against 10% for corpus-linguistically oriented studies (see Section 3.3) and 12% for studies with a computational-linguistic orientation (these will not be discussed here). We will come back later to the minimal involvement of corpus linguists in this sample, but for now, let us note that, while psycholinguists' primary tool of investigation is experimentation, they seem to have taken a liking to corpora and understand the appeal of using the tools of corpus linguists in addition to their own. Interestingly enough, however, psycholinguists appear to use corpora in a way which is quite different from the way corpus linguists use them, as we show presently.

Contrary to corpus linguists, who tend to approach the data in an exploratory fashion, psycholinguists almost always start out with one or more explicitly formulated hypotheses (86% of the psycholinguistic studies in our

sample), which they then go on to test systematically (and usually also statistically). Thus, Monaghan et al. (2005), in their study of the recognition of grammatical categories of words, hypothesize that different cues are useful for different situations, and more precisely, that distributional information is more useful for categorizing higher frequency words, and phonological information more useful for categorizing lower frequency verbs. They then test this hypothesis on the basis of corpus analyses (testing of phonological cues, distributional cues, and combination of phonological and distributional cues) and an artificial language learning experiment.

If we examine the type of phenomenon that psycholinguists are interested in when they use corpus data in combination with experimental data, we notice that syntactic phenomena attract the most attention, with a percentage of 44%. Two topics that seem especially popular are syntactic processing (particularly in combination with subcategorization preferences) and syntactic ambiguity resolution. The former is exemplified by Merlo (1994), who analyzes the influence of frequency (as defined by corpus counts) on the syntactic processing of verb continuations, whereas the latter can be illustrated by Spivey-Knowlton and Sedivy's (1995) study of the online resolution of prepositional phrase attachment ambiguity, which shows that both local information (lexically specific biases) and contextual information (referential presupposition) have a role to play in the process of disambiguation. This type of study can also be found for semantic phenomena, in particular anaphor resolution, as in Long and De Ley (2000), who investigate the effect of the implicit causality of certain verbs on pronoun resolution. The other topics are more evenly distributed than in the corpus-only studies, with 23% of lexical topics, 20% of phonological topics, and 17% of morphological topics. Pragmatics is dealt with in 6% of the papers.

Like corpus linguists, psycholinguists show a preference for written (67%) and specific corpora (59%). There are two main differences with the corpora found in corpus-only studies, however, viz. (i) spoken corpora are almost as frequent as written corpora (61%) and (ii) psycholinguistic studies more often involve the less prototypical corpora discussed above. For example, the corpus exploited by Murfitt and McAllister (2001) consists of monologs and dialogs produced by subjects who were required to describe tangram figures. Poesio et al. (2006) use the TRAINS corpus (Gross et al. 1993), which contains dialogs exchanged in the course of a role-play where one subject played the manager of a railway company seeking to develop a plan in order to achieve a transportation goal, and the other subject played the role of a system providing information such as timetables and equipment availability. It should be added that psycholinguists also use other types of corpus data, which would be more appropriately labeled as 'corpus-derived' data. This includes lexical databases (8%), example collections (3%), and results from other corpus studies (2%).

As is the case in the corpus-only sample, the most common methodology to exploit corpora is frequency (67%). Here, the aim is usually to determine the influence of frequency on cognitive phenomena and the way it correlates with processing. Reali and Christiansen (2007: 4), for instance, combine corpus analysis and self-paced reading experiments "to determine the extent to which the difficulties encountered during online processing of pronominal relative clauses mirror distributional patterns occurring naturally in language". The use of concordance accounts for 42% of the papers from our sample. Collocation clearly plays a subordinate role, as it is used in less than 17% of the studies (to be compared with 32% among corpus linguists, cf. Section 3.1). Sometimes, psycholinguists combine several methods, e.g., concordance and frequency (21%) or collocation and frequency (5%), but our sample does not include any studies which adopt the threefold approach found in corpus-only studies, viz. a combination of concordance, frequency, and collocation.

Crucially, the authors represented in the present sample use some kind of experimentation in addition to the corpus analysis. By combining these two types of data, they also benefit from their respective advantages (see above). Meibauer et al. (2004), who are interested in the use of German *–er*-nominals by children, need to collect experimental data because the available corpora do not provide them with sufficient material to work with. Swerts and van Wijk (2005) are able to identify, on the basis of corpus data, prosodic and lexico-syntactic features coinciding with the use of a particular word order in the Dutch verbal endgroup, but they need an experimental set-up in order to tease these factors apart and establish their unique contribution. Pander Maat and Sanders (2001) exploit yet another advantage of experiments, namely the fact that they make it possible to test the acceptability of fragments not found in corpora (which may be due to a gap in the corpus or to the unacceptability of these fragments). (In studies outside of the sample discussed here, the online nature of experimental data vs. the offline nature of corpora is also frequently put to good use (e.g. Grondelaers et al. 2002 or Gries et al. to appear).)

What is striking in the papers from our sample is the wide diversity of experimental techniques that are employed. What follows is just a short selection of methods found in the sample: primed picture naming (Alario et al. 2004), sentence completion (Bock et al. 2006), semantic similarity judgment (Bybee and Eddington 2006), eye-tracking (Desmet et al. 2006), self-paced reading (Gibson and Schütze 1999), acceptability judgment (Hay 2002), stimulus repetition (Kidd et al. 2007), lexical decision task (McDonald and Shillcock 2001), dictation task (Sandra et al. 1999), vocal imitation (Serkhane et al. 2007). Moreover, it is not unusual for psycholinguists to perform several types of experiments in one and the same study. Thus, in addition to their corpus data, McKoon and Ratcliff (2003) apply as many as five experimental

methods to research the comprehension of reduced relative clauses, namely reading time, acceptability judgment (with measure of the response time), lexical decision, sentence rating, and cued recall.

Finally, we wanted to know what role the corpus data and the experimental data assume with respect to each other in psycholinguistically oriented studies that combine the two types of data. However, this characteristic often turned out to be difficult to establish. In a small number of cases, the authors made it clear that they saw the corpus as a validatee of the experiment (12% of the papers). Gahl (2002), for instance, discusses corpus-derived subcategorization preferences and then performs experiments to test the effect of mismatches of syntactic structures and lexical bias. Similarly, McDonald and Shillcock (2001) develop a measure called *Contextual Distinctiveness* and validate it using lexical decision latencies from their own and other psycholinguists' data. Most of the time, however, this was not sufficiently clear from the paper for us to dare make a decision, and the two methods rather appeared to be on the same footing (89%). Shimojima et al. (2002: 114), for example, explain that they "adopt both an *observational* and an *experimental* approach", thus implying no primacy of one or the other method. Sometimes, the experiment seems to be somehow primary, for example because the corpus data merely serve the purpose of stimulus composition (e.g., Carlson et al. 2005) or because frequency as attested in a corpus is just one of the factors that seek to explain the experimental results (e.g., van Gompel and Majid 2004), but this is not enough to see the corpus as a validatee of the experiment.

3.3. How corpus linguists use experimental approaches

It was already pointed out earlier that, of the 'corpus and experimental' studies from our sample, very few adopted a corpus-linguistic perspective. In fact, this represents only nine papers, two of them published in this journal (Hoffmann 2006; Arppe and Järvikivi 2007) and two (co-)authored by one of the authors of this article (Gries 2003; Gries et al. 2005). Admittedly, this is too small a sample to draw any firm conclusions about how corpus linguists use experimental methods. Yet, some interesting tendencies emerge from this set of studies which are worth mentioning – although they should be treated with all necessary caution. The first general observation that can be made confidently is that, obviously, corpus-linguistic studies that use experimental methods are much less common than psycholinguistic studies that use corpus data (10% vs. 78% in our sample). But it also seems as if corpus linguists combine the two types of data in a different way and with different purposes compared to psycholinguists. Thus, of the nine studies of the sample, only two start with an explicitly formulated hypothesis, six are either exploratory or do not adopt a hypothesis-testing stance, while the final one exhibits characteristics of both these approaches. This is differ-

ent from psycholinguistic studies combining corpus and experimental data, which are predominantly hypothesis-testing, but this is similar to the corpus-only studies.

As far as the topic of investigation is concerned, both syntax and lexis are represented (with three syntactic studies, three lexical studies, and one study combining syntactic and lexical issues; in addition, one study deals with morphology and one with pragmatics). Like the other studies (corpus-only studies and psycholinguistic studies combining corpus data and experimental methods), the data tend to come from a written corpus – but general corpora are more frequent than specific corpora (seven cases of the former and two of the latter). No particular preference is displayed in terms of methodology (equal proportion of concordance, frequency, and collocation). However, as was the case in the corpus-only studies, we notice the use of the threefold methodology concordance-frequency-collocation (Newman and Rice 2004).

Like psycholinguists, corpus linguists who use both corpus and experimental data seem to be aware of the benefits to be gained from such a combination. Lee (2001: 141), for example, notes that elicitation tests allow for "a considerable degree of scientific precision and control" and make it possible to target specific points of interest. However, he also points out that, because of the artificiality of the test situation, they are unlikely to produce natural and spontaneous language, which is precisely what corpora give access to. For Schauer and Adolphs (2006), who are interested in the expression of gratitude, corpus data provide insights into the procedural aspects of thanking someone (including the existence of 'gratitude clusters', which span several conversational turns). On the other hand, the use of elicitation (or, more precisely, a discourse completion task) makes it possible to test the influence of a variety of factors such as formal vs. informal relationship between the interlocutors or high vs. low imposition request. Thráinsson et al. (2007) pinpoint another advantage of elicitation data, namely the fact that, unlike corpora, they enable the linguist to examine what is possible or impossible in a language. By contrast, we could add that corpora reveal what is probable or improbable in a language.

In our small sample of studies, the corpus and the experiment are usually on an equal footing, as was also the case in the psycholinguistic sample. Compared to the latter, however, the experimental techniques employed by corpus linguists show much less variety (which, of course, is partly related to the sizes of the two samples). Of the nine papers, five employ an acceptability/grammaticality judgment task, which can be considered as involving a relatively simple experimental design (as compared, for example, to eye-tracking). While acceptability judgments make it possible to gain insights that could not be provided by corpora and, consequently, can be usefully combined with corpus methods, it should be noted that such data are not

without their problems (see Section 1). Moreover, it is regrettable that corpus linguists seem not to take full advantage of the wide range of experiments that can serve to study linguistic phenomena, as illustrated by the psycholinguistic studies of our sample. We will speculate on reasons for that in the following section.

4. Conclusions and outlook

4.1. Interim summary and conclusions

By examining a sample of recent studies, we have seen that, while corpora represent an important source of information about language, they can also be fruitfully combined with experimental data. However, it is mainly psycholinguists who seem to have realized the potential of such a combination. They usually start out from very precisely formulated hypotheses and integrate the results of corpus studies with a wide range of experimental methods, in order to investigate topics that are of psycholinguistic interest. In terms of linguistic subdisciplines, we discerned a focus on syntactic processing and lexical as well as syntactic ambiguity resolution. Particularly common is the study of the influence of frequency (as attested in corpora) on cognitive phenomena.

Papers with a corpus-linguistic perspective that combine corpus data with experimental methods were rare in our sample. However, it was interesting to note that corpus linguists and psycholinguists seem to approach this combination of data in slightly different ways. More precisely, there were signs that corpus linguists might approach combined corpus and experimental data in roughly the same way as they approach corpus data only. Thus, corpus linguists tend not to begin with rigorous explicit hypotheses before examining the data. They do not hesitate to mix the three main methodologies that are available to analyze corpora, viz. concordance, frequency, and collocation. Finally, they have a particular interest in lexical issues. In corpus-only studies, this is most visible in the choice of the topic, but also, to some extent, in the methodology (cf. use of collocation). This is less clear in the studies combining corpus and experimental data, but the proportion of lexical analyses and collocational methodologies could point in that direction too. This attraction to lexical/phraseological phenomena among corpus linguists who combine corpora and experimentation also appears to be confirmed by other studies not included in our sample but using the same combination of data (e.g., Granger 1998; Källkvist 1998; Schmitt et al. 2004; Gilquin 2007; Siyanova and Schmitt 2008), as well as by the selection of papers brought together in this journal issue (see below).

Given these findings, we believe that corpus linguists should look more into the possibilities of complementing their corpus studies with experimental data. This is relevant for several reasons:

- given the sizes of many currently used corpora, even the smallest results will often be significant and additional experimental evidence will help separate the wheat from the chaff;
- different corpora will yield different results and additional experimental evidence will help us obtain a more precise understanding of phenomena;
- corpus-based results can, and should, be validated against corpus-external findings;
- combining corpus and experimental data would also help us gain insight into the relation between the two types of data.

The first two reasons are straightforward and uncontroversial. The third reason is straightforward but often not recognized. For example, corpus linguists have been developing different quantitative measures of collocational attraction, dispersion of lexical words in corpora, or the importance of words in texts (i.e., keywords). However, there is comparatively little work that attempts to validate, say, the 20+ collocational measures or dispersion measures against findings from corpus-external data and show what, if anything, these measures mean, indicate, or reflect.

The final issue is more complex given the different kinds of findings. Among the studies that use both corpus and experimental data, some have found convergence between them (e.g. Hoffmann 2006), whereas others have found divergence (cf. Roland and Jurafsky [2002], who refer to the inherent differences between "test-tube" sentences and "wild" sentences). This can even be the case when one and the same phenomenon is considered. Thus, Swerts and van Wijk (2005: 246–247) note that, when it comes to the investigation of word order variation in Dutch, some studies have demonstrated a close correspondence between subjects' preferential judgments and their speaking behavior, while others have revealed a huge discrepancy (and sometimes opposition) between the two. Possible explanations have been offered to account for the differences between corpus and experimental data, and suggestions have been made to bring them closer to each other, but it is still true that the relation between the two types of data remains unclear and that identity cannot be taken for granted. Therefore, a study such as Serkhane et al. (2007), where two stages in infants' development are investigated and compared with each other by means of different types of data (one experimental and the other more natural), raises concerns regarding the comparability of the findings.

Finally, let us underline that this combination of corpora and experimental methods should also ideally be integrated with linguistic theory. That such an approach can be very rewarding is demonstrated by McKoon and MacFarland (2000: 856), who see as "the most compelling aspect" of their study "the power gained by combining corpus analysis and psycholinguistic exper-

imentation with linguistic theory, demonstrating how theoretical work and empirical work can inform each other".

The positive picture we are painting here raises the question why so few corpus linguists are using experimental approaches. Very often, the fact that corpus linguists do not also conduct experiments, or at least correlate their data with already published experimental data, seems not to be due to the fact that corpus data are the best or even the only kind of data that allow the question to be studied. Rather, it seems as if part of the reluctance to use experimental approaches or be more concerned with how corpus-based findings relate to more theoretical aspects results from a perceived clash of scientific cultures. At the risk of considerable simplification, a most lively discussion on the Corpora List in August 2008 appears to have shown that corpus linguists can be situated on a cline between two 'extremes'. One extreme views (corpus) linguistics as a humanistic discipline, locates meaning within the text or discourse alone, and treats corpus linguistics and cognitively-inspired approaches as irreconcilable approaches. The opposite extreme views (corpus) linguistics as part of the social sciences, considers meaning, while investigated corpus-linguistically or experimentally, as something that may well be studied outside the text and inside the mind, and regards corpus-based and cognitively-inspired approaches as highly compatible, if not related approaches. We hope, however, that our overview shows that at least converging evidence from different sources should be relevant to everyone working with corpora, regardless of their position on, say, the humanistic-social science continuum, and the papers in this special issue should make this even more obvious.

4.2. The papers in this issue

The papers in this special issue deal with many of the above issues from a variety of perspectives. According to our findings from the literature, it is probably fair to say that the papers do *not* constitute a representative sample of corpus-linguistic studies. Given the focus of this special issue, all papers of course include evidence from corpora and experiments, and the range of experiments spans across much of the kinds of data discussed in Table 1. Some papers use written texts from academic or other genres or from the web (sometimes as their main corpus data, sometimes in order to compare other corpus data with them); some papers rely largely on spoken corpus data. As for experiments, some require the subjects to do things that they do not usually do (produce words when prompted with another word) but which are still not very much out of the ordinary; some require metalinguistic judgments such as judgments regarding the frequency, acceptability, idiomaticity, formulaicity, or value for teaching of expressions – and even within this group, different approaches are adopted (e.g., simple categorical

responses, magnitude estimation, and web-based Likert scales). Then, there is more technical experimentation that involves measuring reaction times before judgments, voice onset times, and duration of pronunciation as well as within-expression priming effects.

However, the present papers also differ from 'typical' corpus studies (as defined by the above overview) in other ways, many of which were included in our desiderata above. For example, most papers adopt a hypothesis-testing approach, relate their findings to theoretical and/or cognitive aspects of linguistic structures, and most involve quantitative methods that sometimes go well beyond mere observed frequencies: *U*-tests, multiple regressions, principal component analysis, generalized linear mixed-effects models, and more are all represented in this special issue and show impressively how sophisticated statistical methods reveal patterns in the data impossible to arrive at by introspection.

The papers by McGee and Nordquist both explore the (mis)match of lexical elicitation and frequencies derived from corpora. McGee is concerned with the degree to which the nouns language teachers provide when prompted to produce the most frequent noun collocates of particular adjectives correspond to collocation data from a general corpus containing both spoken and written data, the British National Corpus. He discusses how his results relate to three different hypotheses, Sinclair's delexicalization hypothesis, Bybee's frequency fusion hypothesis, and Wray's segmentation hypothesis. One particularly interesting aspect of this paper is that the different hypotheses involve statements about the degree to which lexical items are stored, i.e. represented, individually or as part of larger, potentially unanalyzed, expressions.

Nordquist's paper focuses on a very similar question. She, too, compares collocation frequencies from corpus data – this time from a corpus containing only spoken data, the Switchboard corpus – to the first words subjects provided in a cued production task. The hypotheses she tests are the autonomy hypothesis, which states that target collocations are holistically stored and thus relatively autonomous of their component words, and the infrequency hypothesis, according to which even holistically stored collocations may be too infrequent to result in significant lexical effects.

The papers by Wulff as well as Ellis and Simpson-Vlach are concerned with multiword expressions. In the case of Wulff, these have one particular structure, whereas in Ellis and Simpson-Vlach's paper the studied expressions instantiate different syntactic patterns. Wulff studies V-NP idioms exploratorily in the British National Corpus to determine which characteristics are responsible for the feeling that expressions are idiomatic and how these can be defined in a corpus-based way. From a construction grammar perspective, she develops a corpus-driven compositionality statistic and a formal flexibility measure. She first discusses a variety of advantages of this approach and then how it goes beyond previous work, and she shows how different

parameters that correlate with idiomaticity relate to each other and how well they predict idiomaticity judgments by native speakers.

Ellis and Simpson-Vlach study how well statistics that are derived from different spoken and written as well as general and specific corpora and that characterize multiword expressions (most notably Mutual Information [*MI*] statistics and raw frequencies) correlate with various psycholinguistic measures such as formulaicity ratings, response speeds, and judgments. Their paper is particularly interesting because (i) the range of measures involved is quite large, providing a good overview of what corpus measures can and cannot reflect, and (ii) they find that, typically, raw frequencies are mostly outperformed in terms of explanatory or predictive power by a statistic that also takes expected frequencies into consideration, *MI*, something which in spite of many decades of research is still not appreciated by many linguists working with corpus data.

Last but not least, Baroni, Guevara, and Zamparelli study a recent and innovative syntactic pattern in Italian using a very large corpus. Their main point is that the construction in question, the Deverbal Nominal Construction, is in fact a spurious class that is only typical for the particular register of headlinese. They apply the rather recently developed method of generalized linear mixed-effects model to a sample of nearly three thousand examples (from a two-billion-word web-based corpus), and supplement these data with acceptability judgments largely collected via the web. The relevance of this study is particularly grounded in the way in which new statistical methods make it possible to take subject-specific and item-specific characteristics into consideration and develop a data-driven analysis that is actually simpler than what previous, non-data-driven, works would have us believe.

To come back to a previously made comment, the papers here are not particularly representative of contemporary corpus work. However, we hope to have shown that it is exactly their non-representativity that makes this collection of papers so noteworthy. We believe that they – each in their own way – point to exciting new possibilities of using corpus data, possibilities that allow us to extend the range of questions and the score of interpretations and implications considerably beyond what is still the most common kind of corpus-based work. We therefore hope that these papers help to lay the foundation of (i) more interaction between corpus linguists on the one hand and general/theoretical linguists as well as psycholinguists on the other hand and (ii) an endorsement of methodological pluralism, with all the positive consequences that injecting new knowledge and new perspectives has for scientific disciplines.

Acknowledgments

The authors would like to thank Yves Bestgen and Stefanie Wulff for their helpful comments on an earlier version of this paper. Gaëtanelle Gilquin also

wishes to thank Terry Shortall for helping her organize the colloquium "Corpus and Cognition: The relation between natural and experimental language data" at Corpus Linguistics 2007 (Birmingham), where several of the papers included in this special issue were originally presented.

Bionotes

Gaëtanelle Gilquin earned her Ph.D. from the University of Louvain, where she currently works as a researcher for the Belgian National Fund for Scientific Research (FNRS). She is interested in the combination of corpus and experimental data, and more generally in the integration of corpus linguistics and cognitive linguistics. E-mail: Gaetanelle.Gilquin@uclouvain.be

Stefan Th. Gries received his Ph.D. in 2000 from the University of Hamburg. He is currently Associate Professor of Linguistics at the University of California, Santa Barbara. His research interests include quantitative corpus linguistics, statistical methods in linguistics, and cognitively-inspired theories of language. E-mail: stgries@linguistics.ucsb.edu

Notes

1. While motivated in the particular framework under discussion, this methodological change is also curious because one of Chomsky's most important teachers, Zellig S. Harris, broke new ground with his research on empirical/probabilistic approaches to language, which was based on actual texts and is modern in the sense that much contemporary corpus- and computational-linguistic work is still based on his ideas (cf. Goldsmith 2005 for a brief overview).
2. We did not include *Corpora* because of its shorter history (the journal only started in 2006).
3. *Corpus Linguistics and Linguistic Theory* contains a couple of papers combining corpus data with experimental data. These papers were included in our sample of corpus-linguistic studies combining corpora and experimentation (Section 3.3).
4. We do not claim that our coding decisions were easy or are uncontroversial; they just represent our best combined judgment, but they are of course often simplifications and other readers may disagree with regard to particular decisions. For example, the overall perspective of a paper was usually identified on the basis of the content of the paper, but also, in difficult cases, on the basis of the source of the publication and/or the identity of the author(s). It should also be added that the papers were not classified using mutually exclusive categories. For example, a study using more than one corpus method (e.g., concordances and collocations) was counted once for each method, which is why the sums of reported percentages do not add up to 100%.
5. This, in effect, resulted in the selection of relatively recent articles. The vast majority of them (almost 80%) were written in or after 2000, and most of the others in the 1990s. Only three papers date from an earlier period, all of which were discarded from our final sample because they did not fit our selection criteria.

References

Alario, F.-X., Rafael E. Matos, and Juan Segui
2004 Gender congruency effects in picture naming. *Acta Psychologica* 117 (2), 185–204.

Arppe, Antti and Juhani Järvikivi
2007 Every method counts – combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3 (2), 131–160.

Bard, Ellen G., Catherine Sotillo, Anne H. Anderson, Henry S. Thompson, and M. Martin Taylor
1996 The DCIEM Map Task Corpus: Spontaneous dialogue under sleep deprivation and drug treatment. *Speech Communication* 20 (1/2), 71–84.

Bock, Kathryn, Sally Butterfield, Anne Cutler, J. Cooper Cutting, Kathleen M. Eberhard, and Karin R. Humphreys
2006 Number agreement in British and American English: disagreeing to agree collectively. *Language* 82 (1), 64–113.

Borsley, Robert D.
2005 Introduction. *Lingua* 115 (11), 1475–1480.

Branigan, Holly P., Martin J. Pickering, Simon P. Liversedge, Andrew J. Stewart, and Thomas P. Urbach
1995 Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research* 24 (6), 489–506.

Bybee, Joan and David Eddington
2006 A usage-based approach to Spanish verbs of 'becoming'. *Language* 82 (2), 323–355.

Carlson, Rolf, Julia Hirschberg, and Marc Swerts
2005 Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication* 46 (3/4), 326–333.

Chomsky, Noam A.
1964 *Current Issues in Linguistic Theory*. The Hague: Mouton.

Chomsky, Noam A.
1969 Language and philosophy. In Hook, Sidney (ed.), *Language and Philosophy: A Symposium*. New York: New York University Press, 51–94.

Chomsky, Noam A.
1973 Conditions on transformations. In Anderson, Stephen R. and Paul Kiparsky (eds.), *A Festschrift for Morris Halle*. New York: Holt, Rhinehart, and Winston, 232–286.

Connine, Cynthia, Fernanda Ferreira, Charlie Jones, Charles Clifton, and Lyn Frazier
1984 Verb frame preference: Descriptive norms. *Journal of Psycholinguistic Research* 13 (4), 307–319.

Desmet, Timothy, Constantijn De Baecke, Denis Drieghe, Marc Brysbaert, and Wietske Vonk
2006 Relative clause attachment in Dutch: On-line comprehension corresponds to corpus frequencies when lexical variables are taken into account. *Language and Cognitive Processes* 21 (4), 453–485.

Gahl, Susanne
2002 Lexical biases in aphasic sentence comprehension: An experimental and corpus linguistic study. *Aphasiology* 16 (12), 1173–1198.

Garnsey, Susan M., Neal J. Pearlmutter, Elizabeth Myers, and Melanie A. Lotocky
1997 The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37 (1), 58–93.

Gibson, Edward and Carson T. Schütze
1999 Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language* 40 (2), 263–279.
Gilquin, Gaëtanelle
2007 To err is not all. What corpus and elicitation can reveal about the use of collocations by learners. *Zeitschrift für Anglistik und Amerikanistik* 55 (3), 273–291.
Godfrey, John J. and Edward Holliman
1997 *Switchboard-1 Release 2*. Philadelphia: Linguistic Data Consortium.
Goldsmith, John
2005 The legacy of Zellig Harris: Language and information into the 21st century, vol. 1: Philosophy of science, syntax and semantics, Bruce Nevin (ed.). *Language* 81 (3), 719–736.
van Gompel, Roger P. G. and Asifa Majid
2004 Antecedent frequency effects during the processing of pronouns. *Cognition* 90 (3), 255–264.
Granger, Sylviane
1998 Prefabricated patterns in advanced EFL writing: Collocations and formulae. In Cowie, Anthony P. (ed.), *Phraseology. Theory, Analysis, and Applications*. Oxford: Oxford University Press, 145–160.
Granger, Sylviane, Estelle Dagneaux, and Fanny Meunier (eds.)
2002 *The International Corpus of Learner English. Handbook and CD-ROM.* Louvain-la-Neuve: Presses Universitaires de Louvain.
Greenbaum, Sidney
1976 Current usage and the experimenter. *American Speech* 51 (3/4), 163–175.
Gries, Stefan Th.
2003 Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1, 1–27.
Gries, Stefan Th.
2006 Introduction. In Gries, Stefan Th. and Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. Berlin: Mouton de Gruyter, 1–17.
Gries, Stefan Th., Beate Hampe, and Doris Schönefeld
2005 Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16 (4), 635–676.
Gries, Stefan Th., Beate Hampe, and Doris Schönefeld
to appear Converging evidence II: More on the association of verbs and constructions. In Newman, John and Sally Rice (eds.), *Experimental and Empirical Approaches in the Study of Conceptual Structure, Discourse, and Language*. Stanford, CA: CSLI Publications.
Grinder, John T. and Paul M. Postal
1971 Missing antecedents. *Linguistic Inquiry* 2 (3), 269–312.
Grondelaers, Stefan, Marc Brysbaert, Dirk Speelman, and Dirk Geeraerts
2002 'Er' als accessibility marker: On- en offline evidentie voor een procedurele interpretatie van presentatieve zinnen. *Gramma/TTT* 9, 1–22.
Gross, Derek, James F. Allen, and David R. Traum
1993 *The TRAINS 91 Dialogues* (TRAINS Tech. Note No. 92–1). Computer Science Department, University of Rochester, New York.
Harris, Randy Allen
1993 *The Linguistics Wars*. Oxford: Oxford University Press.
Hay, Jennifer
2002 From speech perception to morphology: Affix ordering revisited. *Language* 78 (3), 527–555.

Hoffmann, Thomas
2006 Corpora and introspection as corroborating evidence: The case of preposition placement in English relative clauses. *Corpus Linguistics and Linguistic Theory* 2 (2), 165–195.
Källkvist, Marie
1998 Lexical infelicity in English: The case of nouns and verbs. In Haastrup, Kirsten and Åke Viberg (eds.), *Perspectives on Lexical Acquisition in a Second Language*. Lund: Lund University Press, 149–174.
Kidd, Evan, Silke Brandt, Elena V. M. Lieven, and Michael Tomasello
2007 Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes* 22 (6), 860–897.
Kliegl, Reinhold
2007 Toward a perceptual-span theory of distributed processing in reading: A reply to Rayner, Pollatske, Drieghe, Slattery, and Reichle (2007). *Journal of Experimental Psychology: General* 136 (3), 530–537.
Labov, William
1975 Empirical foundations of linguistic theory. In Austerlitz, Robert (ed.), *The Scope of American Linguistics*. Lisse: The Peter de Ridder Press, 77–133.
Lee, Jackie F. K.
2001 Functions of *need* in Australian English and Hong Kong English. *World Englishes* 20 (2), 133–143.
Lippmann, Richard P.
1997 Speech recognition by machines and humans. *Speech Communication* 22 (1), 1–15.
Long, Debra L. and Logan De Ley
2000 Implicit causality and discourse focus: The interaction of text and reader characteristics in pronoun resolution. *Journal of Memory and Language* 42 (4), 545–570.
Martin, Nadine, Deborah A. Gagnon, Myrna F. Schwartz, Gary S. Dell, and Eleanor M. Saffran
1996 Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes* 11 (3), 257–282.
McDonald, Scott A. and Richard C. Shillcock
2001 Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech* 44 (3), 295–323.
McKoon, Gail and Talke MacFarland
2000 Externally and internally caused change of state verbs. *Language* 76 (4), 833–858.
McKoon, Gail and Roger Ratcliff
2003 Meaning through syntax: Language comprehension and the reduced relative clause construction. *Psychological Review* 110 (3), 490–525.
Meibauer, Jörg, Anja Guttropf, and Carmen Scherer
2004 Dynamic aspects of German *–er*-nominals: A probe into the interrelation of language change and language acquisition. *Linguistics* 42 (1), 155–193.
Merlo, Paola
1994 A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research* 23 (6), 435–457.
Monaghan, Padraic, Nick Chater, and Morten H. Christiansen
2005 The differential role of phonological and distributional cues in grammatical categorisation. *Cognition* 96 (2), 143–182.
Murfitt, Tara and Jan McAllister
2001 The effect of production variables in monolog and dialog on comprehension by novel listeners. *Language and Speech* 44 (3), 325–350.

Newman, John and Sally Rice
2004 Patterns of usage for English SIT, STAND, AND LIE: A cognitively inspired exploration in corpus linguistics. *Cognitive Linguistics* 15 (3), 351–396.
Pander Maat, Henk and Ted Sanders
2001 Subjectivity in causal connectives: An empirical study of language in use. *Cognitive Linguistics* 12 (3), 247–273.
Poesio, Massimo, Patrick Sturt, Ron Artstein, and Ruth Filik
2006 Underspecification and anaphora: Theoretical issues and preliminary evidence. *Discourse Processes* 42 (2), 157–175.
Pullum, Geoffrey K.
1978 Assessing linguistic arguments, Jessica R. Wirth (ed.). *Language* 54 (2), 399–402.
Pullum, Geoffrey K.
2007 Ungrammaticality, rarity, and corpus use. *Corpus Linguistics and Linguistic Theory* 3 (1), 33–47.
Reali, Florencia and Morten H. Christiansen
2007 Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57 (1), 1–23.
Roland, Douglas and Daniel Jurafsky
2002 Verb sense and verb subcategorization probabilities. In Stevenson, Suzanne and Paola Merlo (eds.), *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*. Amsterdam: John Benjamins, 325–346.
Sampson, Geoffrey
1975 *The Form of Language*. London: Weidenfeld and Nicholson.
Sampson, Geoffrey
1980 *Schools of Linguistics*. Stanford: Stanford University Press.
Sandra, Dominiek, Steven Frisson, and Frans Daems
1999 Why simple verb forms can be so difficult to spell: The influence of homophone frequency and distance in Dutch. *Brain and Language* 68 (1), 277–283.
Santini, Marina
2006 Web pages, text types, and linguistic features: Some issues. *ICAME Journal* 30, 67–86.
Schauer, Gila A. and Svenja Adolphs
2006 Expressions of gratitude in corpus and DCT data: Vocabulary, formulaic sequences, and pedagogy. *System* 34, 119–134.
Schmitt, Norbert, Sarah Grandage, and Svenja Adolphs
2004 Are corpus-derived recurrent clusters psycholinguistically valid? In Schmitt, Norbert (ed.), *Formulaic Sequences. Acquisition, Processing and Use*. Amsterdam: John Benjamins, 127–151.
Schütze, Carson T.
1996 *The Empirical Base of Linguistics*. Chicago: The University of Chicago Press.
Serkhane, Jihene E., Jean-Luc Schwartz, Louis-Jean Boë, Barbara L. Davis, and Chris L. Matyear
2007 Infants' vocalizations analyzed with an articulatory model: A preliminary report. *Journal of Phonetics* 35 (3), 321–340.
Shimojima, Atsushi, Yasuhiro Katagiri, Hanae Koiso, and Marc Swerts
2002 Informational and dialogue-coordinating functions of prosodic features of Japanese echoic responses. *Speech Communication* 36 (1/2), 113–132.
Siyanova, Anna and Norbert Schmitt
2008 L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review / La revue canadienne des langues vivantes* 64 (3), 429–458.

Spivey-Knowlton, Michael and Julia Sedivy
1995 Resolving attachment ambiguities with multiple constraints. *Cognition* 55 (3), 227–267.
Stubbs, Michael
1996 *Text and Corpus Analysis: Computer-assisted Studies of Language and Institutions*. Oxford: Blackwell.
Stubbs, Michael
2001 Text, corpora and problems of interpretation: A response to Widdowson. *Applied Linguistics* 22 (2), 149–172.
Swerts, Marc and Carel van Wijk
2005 Prosodic, lexico-syntactic and regional influences on word order in Dutch verbal endgroups. *Journal of Phonetics* 33 (2), 243–262.
Taylor, Charlotte
2008 What is corpus linguistics? What the data says. *ICAME Journal* 32, 179–200.
Teubert, Wolfgang
2005 My version of corpus linguistics. *International Journal of Corpus Linguistics* 10 (1), 1–13.
Thompson, Paul and Alison Sealey
2007 Through children's eyes? Corpus evidence of the features of children's literature. *International Journal of Corpus Linguistics* 12 (1), 1–23.
Thráinsson, Höskuldur, Ásgrímur Angantýsson, Ásta Svavarsdóttir, Thórhallur Eythórsson, and Jóhannes Gísli Jónsson
2007 The Icelandic (pilot) project in ScanDiaSyn. *Nordlyd* 34 (1), 87–124.
Tummers, José, Kris Heylen, and Dirk Geeraerts
2005 Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory* 1 (2), 225–261.
Wasow, Thomas and Jennifer Arnold
2005 Intuitions in linguistic argumentation. *Lingua* 115 (11), 1481–1496.