Corpus-driven Cognitive Semantics
Introduction to the field

Dylan Glynn

Is quantitative empirical research possible for the study of semantics?[1] More specifically, can we use corpus data to produce testable and falsifiable results for the description of meaning? If so, what are the advantages and what are the limitations of such an approach? Such questions are often heard and it would seem they are behind much of the reticence and scepticism for corpus-driven techniques in the broader Cognitive Linguistic community. This volume grew out of the workshop Usage-Based and Quantitative Methodology in Cognitive Semantics at the International Cognitive Linguistics Conference 2007, in Kraków, where the participants were asked to answer these very questions. This book both critiques and supports the application of such techniques to the study of meaning in language.

1.   Cognitive Linguistics – A necessarily empirical approach

Cognitive Linguistics is beginning to realise the implications of its own theoretical framework. When Langacker (1987, 1988) first outlined the theory of entrenchment and his usage-based model, he may not have appreciated the extent to which these proposals would be implied in the trends of 21st century linguistics. Indeed, Langacker's theoretical proposal arguably set the stage for a major paradigm shift in linguistics, from theory-driven to empirical research. Similarly, when Lakoff (1982, 1987) argued that we need to approach language holistically, that categorisation based on human experience is the foundational structure of meaning, he is likely to have underestimated the implications of this move. A holistic approach to meaning goes far beyond adding pragmatics to semantic analysis – it entails that everything speakers know of the world, including the whole domain of sociolinguistics, is necessarily drawn into the mainstream of language description.

    It is due to these theoretical tenets that the roots of Cognitive Linguistics were, and inevitably so, an empirical movement seeking to account for the

totality of language as a socio-cognitive phenomenon. The non-empirical emphasis of earlier research can be attributed to methodological practices inherited from the Structuralist and Generative (or Mentalist) schools. These schools assume an object of study based on society's *langue* or the ideal speaker's competence, and as such, their methods can rely exclusively on introspection.

Cognitive Linguistics' model of language necessarily entails empirical methodology. In fact, Fillmore (1985), Talmy (1985), Lakoff (1987), and Langacker (1987) can be said to have produced one of the first data-oriented models of language. Their model dismantles a hundred years of fervent linguistic theorising. They deconstruct the Structuralist distinction of semantics and pragmatics, demonstrate the arbitrariness of the Modularist division between lexis and syntax, and argue that language structure is a result of usage, based on general cognitive capacities. The object of study is finally language use in all its complexity, rather than *langue* or 'competence', lexis or syntax, and semantics or pragmatics.

The model of language proposed by Cognitive Linguistics is so completely simple that it places the emphasis squarely on method and data. Rather than simplifying the object of study by carving off its complexities with hypothetical modules of language structure, it lands the linguist in the midst of a chaotic phenomenon that is the nature of all socially structured systems. The original cognitive theoreticians did this in dialectic with Structuralism and Mentalism. Most linguists today would agree that the Cognitivists succeeded in demonstrating how such reductionist theories failed. But did the Cognitivists offer a viable alternative? In some sense, they took us out of the frying pan and into the fire. Firstly, claiming that grammar is semantics-driven and that linguistic meaning extends beyond traditional semantics to all world knowledge exponentially increases the complexity of the object of study. Secondly, both Structuralism and Mentalism claimed that early Cognitive Linguistics offered no means for testing hypotheses or falsifying analyses. They were right. Lakoff's (1987) claim that co-occurrence of semantic phenomena is the basis of Cognitive Linguistics' scientific method is largely vacuous unless results obtained with this method can be falsified. Likewise, it is all very well for Langacker (1987) to argue that grammaticality is relative and varies from person to person. But if this is the case, how can we test proposed grammatical structures?

Despite the fact that neither Lakoff nor Langacker developed an empirical line of research, together their proposals establish the cornerstone of

true empirical linguistic enquiry. Lakoff argued that evidence for his proposals comes from the co-occurrence of linguistic phenomena. Reframed, his argument was for inductive analysis, a method of analysis which is the norm in the social sciences. Making generalisations based on a sample, then extrapolating those generalisations to the population is the basis of inductive scientific research, and the only viable method for the social sciences. Yet, without rigorous and testable techniques for establishing how representative or reliable a generalisation is, the 'results' remain merely hypotheses based on very small samples. Indeed, as was typical at the time, a sample was no more than the internal language knowledge of the linguist and, perhaps, a few colleagues.

It is here that Langacker's theory of entrenchment and the usage-based model step in – they offer an operationalisation of grammar and grammaticality. In empirical research, operationalisation is a basic analytical tool – the definition of an object of study by how one can measure it. Once dealing with a measurable object of study, one may test the accuracy of its description and falsify predictions made about it. The ability to falsify results is crucial if language analysis is to progress without grammaticality tests. Langacker's theory allows us, therefore, to test the accuracy of generalisations based on Lakoff's identification of co-occurring linguistic phenomena. These two points are crucial for quantitative corpus-driven semantic research and we return to these points below.

Having established that the theory of Cognitive Linguistics is inherently empirical, can we say, however, that the paradigm is empirical in practice? To some extent, empiricism, be that based upon the experimental data of psycholinguistics or upon the collected data of corpus linguistics, is a new turn for Cognitive Linguistics. Yet, despite the recent expansion of empirical methods, both corpus-driven and experimental research go back to the origins of Cognitive Linguistics.[2]

Since the very beginning of the movement, a small yet important number of scientists have taken very seriously the methodological implications of the cognitive linguistic model of language. Two early works, by scholars whose names were crucial in the foundational years of the paradigm, are testimony to this. Firstly, Dirven *et al*. (1982) is a corpus-driven study of depth and maturity that would hold its own today. Secondly, Lehrer (1982) is a linguistic experiment whose design and thoroughness exceed many of our current studies. The former investigation led, directly and indirectly, to the development of corpus Cognitive Linguistics in Belgium and Germany. The latter work, recently republished by Oxford University Press (Lehrer

2009), is enjoying renewed attention a quarter of a century later. Although the experimental track is equally important to the development of Cognitive Linguistics, it lies beyond the purview of this volume.[3] The current volume belongs to the tradition, established with Dirven *et al*. (1982), focusing on methods employing corpus data.

From Dirven *et al*. (1982), via Dirven & Taylor (1988), Schulze (1988), Rudzka-Ostyn (1989, 1995, 1996) and Goossens (1990) up until Geeraerts (1993b), Schmid (1993, 1997, 2000) and Geeraerts *et al*. (1994, 1999), corpus-driven Cognitive Linguistics grew slowly but surely. During the 1990s, the movement was given added weight by the demonstration of the inadequacies of the analytical techniques current at the time. The Structuralist and Mentalist-Modularist assumptions about how language works, assumptions that still drove the analytical techniques of early Cognitive Linguists, were shown to be ineffective. Geeraerts (1993a) demonstrated theoretically that the study of polysemy still made Structuralist assumptions about how meaning is organised. Essentially, the kind of radial network analysis popularised by Lakoff (1987) sought to identify discrete lexical senses where there was no reason to suppose that such discrete senses exist:

> The tremendous flexibility that we observe in lexical semantics suggests a procedural (or perhaps 'processual') rather than a reified conception of meaning; instead of meanings as things, meaning as a process of sense creation would seem to become our primary focus of attention. Geeraerts (1993a: 260)

Two years later, Sandra & Rice's (1995) experimental study 'Mirroring whose mind, the language user's or the linguist's demonstrated that the techniques used to study polysemy in Cognitive Semantics were, in Popper's sense, *ad hoc*. Just as the approaches in Cognitive Linguistics had inherited the analytical construct of Structuralism - senses come in discrete categories - they had also inherited that the analytical technique of the Mentalist grammars – introspection can identify grammaticality. However, for Cognitive Linguistics, with no 'ideal speaker', the intuition of one native speaker (the linguist) cannot be assumed to represent the language. Introspection is still widely used in Cognitive Linguistics, but following the publication of Sandra & Rice's study, descriptive semantic analysis based entirely on introspection, lost much scientific credibility within the cognitive community.

This critique of introspection as the sole basis of linguistic investigation, in combination with the establishment of Cognitive Linguistics as an independent model of language, led to the growth in experimental and corpus-

driven methods. To understand, however, the extent to which Cognitive Linguistics has always been a empirical approach to language, let us briefly revisit the foundational theories of Fillmore (1985), Talmy (1985), Langacker (1987), and Lakoff (1987).

2.   Usage-based Cognitive Linguistics - unifying Mentalist and Structuralist theories, operationalising meaning and grammar

Geeraerts (2006a: 398-415, 2006b, 2007, 2010) argues that Cognitive Linguistics has recontextualised language description, returning elements that were removed by Structuralism and Mentalism-Modularism. More specifically, he shows that the tenets of the paradigm entail that meaning drives grammar. This meaning, he stresses, is conceptual, emergent, and social. Geeraerts' idea of the socio-conceptual recontextualisation of Cognitive Linguistics is difficult to dispute, seeing the current trends. The basic argument is that there has been a linear progression from Structuralist and then Mentalist-Modularist decontextualisation to the recontextualisation of grammar in Cognitive Linguistics. However, one could perhaps go further and argue that this is but a symptom of a more fundamental unification of language theories. There are two issues here - how the theory of Cognitive Linguistics unites Mentalist and Structuralist language models, and how this unification operationalises grammar and semantics. As such, it can be said that a combination of the usage-based model and corpus-driven research resolves some of the great debates of 20th century linguistics.

2.1. Entrenchment – operationalised grammar

Langacker's (1987) theory of entrenchment unites the two dominant 20th century models of language. If we read Langacker's position in Chomskyan terms, the idea of an individual's 'competence' is maintained.[4] Each speaker possesses a mental grammar with intuitions about correctness in language. However, for Langacker, this individual mental grammar does not represent a language. In other words, there is no ideal speaker - each individual's grammar is a learnt code that is unique to the extent that the individual's exposure to language is different from other individuals' in the speech community. It is this point that is crucial since it brings us to de Saussure's *langue*. The speaker, in his or her use of language, constantly

judges a perceived set of usage norms, which in Structuralist times, would have been called *langue*.

Entrenchment, however, is more than a theory of grammaticality, it is an operationalisation of grammaticality. Langacker is primarily concerned with the status of a linguistic unit, but in the usage-based model, entrenchment can be extended to replace the notion of grammatical correctness, where the principle of frequency of use for the individual is extended to that of frequency of occurrence in the community. This operationalisation defines the phenomenon of grammar by showing how one may observe and measure it.

> Every use of a structure has a positive impact on its degree of entrenchment, where […] disuse has a negative impact. […] Moreover, units are variably entrenched depending on the frequency of their occurrence (*driven*, for example, is more entrenched than *thriven*)… The absence of a sharp division between units and nonunits has the consequence that the scope of a grammar is not precisely delimited. (Langacker 1987: 59-60).

Here, Langacker offers an observable and quantifiable definition of an individual's grammar and demonstrates how it relates to the grammar of a language.[5]

The measurable definition of grammar proposed by Langacker unifies the internal Mentalist and the external Structuralist conceptions of language. This unification eliminates the need for both the *langue-parole* and competence-performance distinctions. It restores the holistic complexity of language analysis by treating the emergent structure of grammar as a result of its use by individuals in a community. Seen in this light, not only is Cognitive Linguistics recontextualising language as a social phenomenon, it is uniting the two principal 20th century models of language and defining its framework in such a way as to permit empirical research.

2.2. Categorisation - operationalised encyclopaedic semantics

Entrenchment, as a theory, bridges Mentalism and Structuralism. It is, however, a symbolic theory of grammar, based on the pairing of form and meaning. Thus, semantics is crucial and basic to any analysis that employs it. Can we also operationalise meaning? Stefanowitsch (this vol.) asks this very question and offers four operational definitions. However, before we ask that question, we must emphasise that meaning is not restricted to propositional or referential semantics in Cognitive Linguistics. Lakoff's

(1987) theoretical work represents the foundations of Cognitive Semantics and his position mirrors Langacker's unification of *langue* and 'competence' in that it also dismantles hypothetical sub-parts of language structure, such as linguistic semantics and pragmatics. Following Fillmore's (1985) work, Lakoff argues for a non-propositional semantics based on world knowledge – encyclopaedic semantics.

Yet, encyclopaedic semantics should not be seen as merely a combination of linguistic semantics and pragmatics; it places the holistic complexity of human experience and semiosis in the driving seat of language production. The study of language must, therefore, account for all that we know about the world from denotational reference, through cultural knowledge to even social expectations. In concrete terms, any study must simultaneously account for the semantic motivation behind and interaction between syntax, morphology, lexis, prosody, and all of this relative to discourse structures, world knowledge, and social variation. Moreover, this sea of infinite complexity varies subtly from individual to individual. It is the sum of all this that represents our object of study – meaning. The seeming impossibility of scientifically describing such multidimensional complexity is what led Structuralism to treat language use separately, the *parole*, and what led Mentalism-Modularism to dismiss it entirely. However, for a non-modular semantically driven theory of language, the complexity of language use is the basis of grammar. Can such a broad-reaching understanding of meaning be operationalised scientifically? Lakoff attempts this through the notion of conceptual categorisation.

Categorisation is a symbolic distinction between difference and similarity. This notion permits Lakoff (1987) to operationalise meaning. Things that are conceived as similar are grouped together, distinct from things that are not similar. The sets of things conceived of as similar are called concepts. This use of the notion concept is a powerful operationalisation of meaning and allows us, as linguists, to use the same analytical tool to account for the full complexity of encyclopaedic semantics. It explains why two physically different chairs are both labelled *chair*, why the lexeme *bill* is not ambiguous in a restaurant scenario, why *she sneezed the napkin off the table* is comprehensible, but also how conceptual metonymy and metaphor function.

The idea that, as cognitive beings, we are constantly judging similarity and dissimilarity, that through this ability we order the chaotic and dynamic environment in which we live, and that this ability and process is basic to language seems reasonable. Moreover, at a theoretical level, it is a reason-

able operationalisation of meaning. But does it work? Can this abstract definition lead to quantification and measurement? In fact, it is already the basis of much corpus-driven research. Co-occurrence and correrlation are fundamental to corpus investigation. Whether it is formal co-occurrence, such as that indentified by collocation in Collostructional Analysis, or semantic co-occurrence, such as that indentified in the usage-feature analysis of the behavioural profile approach, corpus research functions by identifying sets of similar things. This is exactly Lakoff's understanding of conceptual categorisation. So we see that just as frequency can operationalise grammaticality, co-occurrence can operationalise categorisation.

Thus, we can say that frequency of co-occurrence, which is fundamental to corpus research, is a quantitative operationalisation of the basic theories of Cognitive Linguistics – entrenchment and categorisation. These theories, entrenchment and categorisation, explain grammar and meaning. However, co-occurrence must be understood as much more than formal co-occurrences. As stressed above, we must simultaneously account for the interaction of all dimensions of meaning in order to explain language. But it is precisely this daunting task that quantitative corpus-linguistics is best placed to achieve. Let us consider how.

2.3. Linguistic complexity - A multifactorial approach to language

We have so far established that Cognitive Linguistics has freed linguistic research of complex theoretical models, the models that were designed to tie down and render the complexity of natural language simple enough to study rigorously. We have also seen how grammar and meaning are accounted for theoretically and how these accounts are operationalised in terms that are applicable to corpus-driven research. Yet still, how is it possible to scientifically account for such an immense and complex linguistic system, a system that varies from speaker to speaker and from context to context? With no core-grammar, no *langue*, not even propositional or referential semantics, what hope do we have of capturing, accurately and rigorously, the conceptual structure that we believe motivates language? Is it truly possible to make generalisations and write grammars without reducing this complexity? One answer lies in multifactorial modelling. Of all the advantages of corpus-driven research, this stands out as one of the most important. Geeraerts (2006a) summarises succinctly the basic advantage:

> Corpus materials provide the firm empirical basis for research, in which lin-
> guistic phenomena are statistically analysed with the help of multivariate
> techniques, and in which social and cultural variation is explicitly included
> into the multifactorial model. Geeraerts (2006a: vi)

Multivariate statistics is a powerful tool for any social scientist. Quantita-
tive methods are essential in determining the probability that our generali-
sations are representative of the population but also because they help find
patterns of co-occurrence that might be otherwise impossible to identify.
This is especially the case when faced with so complex and multidimen-
sional a phenomenon as language. Without language modules such as se-
mantics, pragmatics, lexis, or syntax, language description must account for
the simultaneous interaction of the different dimensions of language. In-
deed, not only is multifactorial modelling of language, and the various fac-
tors that go into speech, necessary to capture the said complexity, it also
represents a cognitively plausible model of language production.

   Multifactorial analysis began by operationalising the various factors that
impact upon the use of lexemes, morphemes, and syntactic patterns. This
was done through feature analysis of large numbers of found examples.
Despite the success of these first attempts (Dirven *et al.* 1982; Schmid
1993, Geeraerts *et al.* 1994; Rudzka-Ostyn 1995), manually handling the
results of multifactorial feature analysis limits their interpretative power.
Inspired by the research of psycholinguistics and sociolinguistics, Geer-
aerts *et al.* (1999), Gries (1999, 2001, 2003), and Speelman *et al.* (2003)
began applying multivariate statistics to language analysis.

   With the development of quantitative tools to treat corpus data and el-
icited data, the future looks bright. It is not difficult to imagine a future
where linguists learn different analytical methods rather than study hypo-
thetical constructs. In such a scenario, with the old dialectic between Cog-
nitive Linguistics and its Mentalist and Structuralist precursors forgotten,
linguists would no longer need to identify themselves as one theoretical
camp or another. Rather, on the basis of empirical and verifiable data, lin-
guists would dispute which method of data collection and analysis is best
suited to answer a given research question. Similarly, they would question
the representativity of data instead of simply choosing a different example
that better suits their hypothesis. Linguists would compare results gleaned
from various methods, and only then, equipped with tested hypotheses,
would they advance theoretical debate. As Geeraerts (this vol.) stresses,
differences in results that either confirm or falsify hypotheses, should be
the only basis for theoretical discussion.

3.   Quantitative corpus-driven Cognitive Semantics

Employing the usage-based model means that real language use, in all its complexity, must be the basis of linguistic research. But what does usage-based methodology entail? From corpus-extracted illustrations of a theoretical discussion to collocation studies and then confirmatory multivariate corpus-driven analysis, there is a full gamut of methodological variation (cf. Tummers, Heylen & Geeraerts 2005 and Heylen, Tummers & Geeraerts 2008 for a discussion on this point; Gilquin & Gries 2009 extend this discussion to the use of experimental data). Before quantitative corpus-driven Cognitive Semantics can become a mainstream methodological choice for linguists, it must answer two very important questions. How can we use numbers to analyse semantics and why should we do so? The problem is not corpus data or even corpus-driven analysis. Deignan (2005, 2009), for example, offers excellent examples of non-quantitative corpus-driven Cognitive Semantic research. The problematic issue is quantification. Meaning is an inherently subjective, mentally internal, and non-observable phenomenon. Why should we attempt to analyse it with quantitative techniques?

Before we can consider the strengths and weaknesses of quantitative Cognitive Semantics, we must set straight two common misconceptions about corpus methods that persist within the wider Cognitive Linguistic community.

3.1. Fallacies about corpus methodology in Cognitive Linguistics

Many linguists believe that corpus-driven research is restricted to the study of collocations and that quantitative techniques used to treat corpus data are restricted to obtaining probability values. These two beliefs are erroneous. Firstly, corpus linguistics is not restricted to the study of collocations and formal correlations. Although the study of 'words and the company they keep' lies at the origins of corpus linguistics, methods for studying corpus data have developed much since those times, both in terms of the kinds of questions that are asked and the kinds of techniques used to answer them. This is not to say that collocations and syntactic patterns are no longer important within the field, as we will see in the following chapters, but they are far from the only research paths possible using corpus techniques.

Secondly, quantitative techniques are not just about *p*-values. The belief that corpus linguistics is all about probability scores and that corpus linguists do not 'get their hands dirty' with careful close language analysis could not be further from the truth. Moreover, this is particularly false when corpus techniques are applied to semantic research. Indeed, the usage-feature or behavioural profile approach, one of the principal trends in corpus-driven lexical semantic research (Gries 2006a; Divjak 2006; Glynn 2009; Janda & Solovyev 2009; Speelman & Geeraerts 2010) and grammatical semantic research (Heylen 2005; Tummers *et al*. 2005; Wulff *et al*. 2007; Grondelaers *et al*. 2008; Szmrecsanyi 2010) involves the detailed and extremely laborious task of analysing a wide range of formal, semantic, and sociolinguistic features of thousands of natural language examples. Given this kind of research, arguing that corpus linguists are just interested in numbers is risible.

The quantitative treatment of data made available through meticulous manual or semi-automatic analysis, should not be seen as the hunt for *p*-values. Statistics has several roles – estimating statistical signficance; detecting patterns in data; identifying the relative importance of usage factors in grammatical composition or word choice; and determining whether a proposed model or explanation accurately describes the data at hand. Probability values are but one of these roles, they indicate statistical signficance. They are important because they tell us that our results would most likely be replicable, but they say no more than that. Save as confirmation that a given analysis and its interpretation are more than chance, *p*-values offer no linguistic insights *per se*.

If we accept that corpus linguistics can be more than the study of formal phenomena and that statistics is more than the search for *p*-values, then we can move on to consider the true strengths and weaknesses of quantitative corpus-driven Cognitive Semantics. These questions are left up to the authors of this volume, but let us briefly consider the main issues that lie behind many of the debates in the field.

3.2. Issues for corpus-driven semantics

3.2.1. Corpus representativity

There is, in reality, no such thing as a balanced corpus and no corpus can ever hope to be representative of a language. Corpus linguists have argued

that despite this shortcoming, a corpus is surely more representative than a single speaker. This response is not as unproblematic as it sounds, raising fundamental questions about our object of study. Modern linguistics was founded by de Saussure who delimited the object of study, *langue*, as a structured system in a given place and time. His ideas set the stage for the erroneous assumption that language is a discrete object of study, whose delimitation is not an issue. In some ways, much of the spilt ink of 20th century linguistic theory directly follows from disagreement over the object of study. From syntax and core grammar to pragmatics, linguistics argued over what language is for most of the preceding century. As stressed above, Cognitive Linguistics unifies modular language analysis with a holistic recontextualised approach to language. In light of this, one might argue that a corpus, which cannot represent the entire complexity of a language, can never be a basis for studies of language, writ large. The response is straight-forward - we do not attempt to account for all of language in every study.

The usage-based model places variation, between groups and even be-tween individuals, as an integral part of language. This we can accept *a priori*. The implications of such an assumption are brought forward when we build our corpora or extract our data for analysis. Indeed, in each cor-pus-driven study, the linguist is forced to answer the question: what part of language are we studying? Often practical issues dictate such choices, but in any case, they are always overtly recognised choices. These choices di-rectly determine the scope of a study and this information should reflect variation inherent in language. Nevertheless, fine-tuning our research methods and building better, more diverse, corpora remain important keys to improving representativity. Gries (2006b; 2008) examines some of the issues at hand and the general push towards more analyses of spoken cor-pora is well documented in corpus linguistic journals (cf. Newman 2008a and this vol.).

It is impossible to study all of language at once (the reason why the Structuralist and Modularist programmes chose to focus on 'parts' of lan-guage). Corpus linguistics makes this impossibility overt and, as such, this apparent limitation is actually a blessing in disguise. It will help linguists be much more realistic about what can be scientifically said about language in a single study.

3.2.2. Negative evidence

One of the most commonly cited criticisms of corpus linguistics, a criticism that goes back to Chomsky, is that there is no negative evidence. No corpus, irrespective of its size, can possibly represent a language, let alone tell us whether a given expression or use of an expression is impossible. However, negative evidence in this sense of the term is only crucial in a rule-driven theory of language, such as that propounded by Chomsky. Since, in a usage-based model, grammaticality is based on entrenchment, and entrenchment is a result of use varying from individual to individual (indeed, even within the individual), there are no hard and fast rules. In a usage-based model of grammar, grammatical rules are merely generalisations about usage. For such a model of language, negative evidence is of much less importance.

To take a simple example, in a Mentalist rule-driven grammar, for the proposal of a rule that the second person copula in English takes the form *are*, any occurrence of **you is* or **you am* would disprove the rule. However, since grammaticality is a matter of degree in Cognitive Linguistics, the fact that in certain regions of England certain social groups say *you is* (especially in the reduced form) is incorporated into our grammar, without negating the constructional importance of the *you are* pattern. Instead, the use of this construction is relative to established sociolinguistic factors and is part of the system. Hence, grammatical tests and their need for negative evidence are of no concern to the usage-based linguist. Langacker's operationalisation of grammaticality means that frequent is grammatical and infrequent is less grammatical.

For Structuralism, defining the meaning of a lexeme, such as bachelor, might be an example of the need for negative evidence. If we define a bachelor as an unmarried man, but there exists a propositional use that contradicts this meaning, such as *??a Catholic priest is a bachelor*, the definition would be shown to be inadequate. In this case, one would need to redefine the lexeme (here as a "man thought of as someone who could marry" Wierzbicka 1990). The process is methodologically identical to the rule (re-)writing of Generative Grammar. If a proposed rule generates sentences that are unacceptable, then the rule is too powerful and needs to be redefined. In this way, the desire to identify the necessary and sufficient conditions for distinguishing the meaning of one lexeme from the meaning of another motivates the need for negative evidence. However, the idea of a proof, which may be falsified with a single counter example, is arguably

not informative for language description. Rather, social sciences are better served by generalisations based on samples. Given the model of language, the need for negative evidence is not so obvious for the cognitive linguist. For further discussion on this issue, within the framework of corpus Cognitive Linguistics, see Stefanowitsch (2006, 2008).

3.2.3. Frequency and salience

What can frequency tell us about language? This is a very real issue for the corpus-driven study of meaning. Although some might argue that frequency is an indicator of productivity or entrenchment, it would be difficult to claim that frequency directly equates salience.[6] If salience is an indicator of relative semantic importance, we are faced with a problem. In terms of semantic content, most frequent often equates least semantically important, where rarity, or marked usage, indicates greater semantic importance. Givón (1991) is explicit on this point - "The marked category [...] tends to be less frequent, thus cognitively more salient, than the corresponding unmarked one." However, in another sense of semantic importance, the reverse is sometimes true. For example, when a language more frequently refers to a given phenomenon, cultural or otherwise, we assume that it is culturally significant. This line of reasoning is basic to much of Wierzbicka (1985) and Lakoff's (1987) research on culturally determined categories.

Salience is often cited as a crucial notion for establishing relations between categories, especially prototype structures (Durkin & Manning 1989, Geeraerts 2000). It follows that since the results of psycholinguistics are often based on relative categorisation determined by salience, understanding the relationship between frequency and salience is essential for the comparison of corpus and experimental results. Thus, although there is a relationship between frequency, salience, and semantic import, this relationship is not straightforward. Since corpus research is dependent on the study of relative frequency, how can this method be used to talk about semantic structure and how can its results be compared to experimental results?

This question may be answered by operationalising the concepts involved. It appears that there are different kinds of salience, yet how frequency is related to perceptual salience, conceptual cultural salience, or formal linguistic salience is still an open question. This debate is growing within the field and is treated by Newman (this vol.) and Schmid (this vol.),

but see also Gilquin (2006, 2010), Wiechmann (2008a, 2008b) Gilquin & Gries (2009), and Arppe *et al*. (2010). The answers to such questions will probably be found when concepts such as salience are more clearly defined through operationalisation.[7] Perhaps, it will be shown that frequency can be a measure of certain kinds of salience and not others. Currently, however, the debate remains open.

3.3. Advantages of corpus-driven Cognitive Semantics

This volume should not only be seen as an attempt at identifying the limitations of a corpus-driven method, the advantages must also be developed. Let us briefly consider the basic advantages of this approach.

3.3.1. Empirical

It may seem obvious that corpus-driven research is empirical, but it is important to state this clearly because, for two different reasons, many within the research community believe that the use of corpus data does not constitute empirical research in linguistics. The first argument comes from within the experimental community and is discussed by Gries & Divjak (this vol.). Summarised briefly, the argument is that as corpus linguists, we do not have direct access to the mind, and therefore language production. It follows that we are not actually studying language production, but traces of language use extant in corpora. The response to this critique is simple - it is true. No corpus linguist studies language production. However, these traces of use are a wonderfully rich source of information on how language is used. Moreover, although we have no direct access to our object of study, nor does experimentation. Psycholinguists elicit responses to stimuli; they do not look inside the 'black box' itself. Thus, we collect data, they elicit it. Neither approach has direct access to the mind or its functioning. Psycholinguistic experimentation is better placed to look at processing and dynamic issues in language production, while corpus linguistics is better placed to look at natural use and, perhaps, even the intersubjective dimension of language (cf. Glynn & Krawczak submitted). Although, as stressed by Newman (2008a; this vol.), there is a terrible lack of spoken and interactive corpora for exploring these dimensions of language.

The second argument comes from the community that uses introspection to perform linguistic analysis. Their argument is reasonable - since our theory of language is a semantics-driven theory of language and semantics is necessarily subjective, that is internal to the mind, introspection is the only viable method for its analysis. Talmy (2008: xix) is explicit about this point. He stresses that corpus research "cannot directly yield many abstract linguistic patterns". His argument is similar to that of the psycholinguists', save, at least, that he is correct in saying that introspection does access the mind (unlike in psycholinguistic experimentation). His point is not disputed but we have two important additions to the discussion. Firstly, despite this limitation, the patterns of natural language usage are an incredibly rich source for working out how people use language and this can produce a very good picture of language structure. Secondly, as stressed by Geeraerts (this vol.), introspection is vital to corpus linguistics. The categories chosen for study, the actual analysis of those categories, and, of course, the hypotheses that the study will test are all a result of introspection. It is not that empirical research replaces introspection, rather introspection is used to propose hypotheses, which then need to be tested by operationalising the questions and designing a study that will adequately answer those questions. If we cannot find a means to adequately answer a scientific question, then we are in the same position as scientists in all fields and must continue to look for a means to do so. This is no argument to remain at the hypothetical stage of enquiry, based entirely upon introspection.

3.3.2. Quantitative

If one has empirical data, then one may quantify them. However, it is far from self-evident that corpus-driven semantic research should be quantitative. Since debating this proposal is the point of the current volume, we will not delve too deeply into this discussion. However, the basic advantage offered by a quantitative approach to semantics is that it permits the operationalisation of our studies. The benefits of, and indeed need for, operationalisation are discussed by Stefanowitsch (this vol.). The questions, of course, are – can and should we operationalise semantics quantitatively? These questions are specifically broached by Newman (this vol.) and Schmid (this vol.) but, to a greater or lesser extent, they are considered by all the contributions to this volume. One basic point should be established - no one wishes to argue that we should reduce semantic questions to num-

bers. It is not the frequency *per se* of linguistic features that is of interest, but what this says about usage, the relative association of forms and meanings in context.

Even if quantification is possible, the next question is - why do we need it, what do we gain from employing quantitative techniques? Firstly, they permit, and indeed encourage, the empirical cycle, explained by Geeraerts (this vol.). It is much easier for the research community to verify the results of a study when one can see exactly how those results were obtained. Secondly, quantified data can be examined using statistics. Statistical analysis of data offers several advantages: (i) Confirmatory statistics allows one to determine the statistical significance of the results of an analysis. In other words, what is the probability that similar results would be obtained if further examples were examined in the same way? (ii) Multivariate statistics allows one to identify patterns in usage that would be effectively impossible to identify using introspection. (iii) Statistical modelling allows one to test the accuracy of an analysis. This kind of measure determines how much of the variation, in a given sample, a given analysis can explain. These multivariate techniques do not only examine the effects of numerous factors of usage simultaneously, they include the interaction of these factors as well. Thus, Empiricism permits quantification and quantification permits statistical confirmation but also multifactorial analysis. This is the third basic advantage.

3.3.3. Multifactorial

Empirical research permits quantification, and, in turn, quantification permits multifactorial analysis. Although this may seem like a narrow addition to the discussion, it is arguably the most important facet of corpus-driven semantic research. As stressed in section 2.3, it is the complexity of our object of study that is our greatest hurdle. The cognitive model of language insists that we simultaneously take into account the full socio-cognitive spectrum of factors that influence and motivate language. Whether the research is based on questionnaires, advanced eye-tracking experiments, word space models derived from mass computation of huge corpora, or the nitty-gritty manual analysis of semantic features of a small corpus sample, multivariate modelling is a basic and crucial tool. It is almost impossible to imagine how one might account for all the factors, and the interaction of these factors, that go into language production and comprehension. How-

ever, should it ever be possible, it will only be so with the aid of such ana-
lytical tools. Therefore, arguably, it is precisely the possibility of multifac-
torial analysis that makes quantitative empirical research, be that corpus-
driven or experimental, essential for linguistic semantics.

4.   Cognitive Semantics – lexical and grammatical meaning

This decade has seen a veritable surge of corpus-driven Cognitive Lin-
guistic research. This volume can only scratch the surface of the theoretical
questions that such a method raises and offer a small sample of the research
it makes possible. The focus of the volume is specifically the application of
quantitative corpus-driven methods to Cognitive Semantics. Yet, even lim-
iting the horizons in this manner, the possible applications and issues go
beyond the purview of any single volume. Let us place the work presented
here in its methodological context. This will allow the reader to better
understand its importance, but also see where future directions of the field
lie.

4.1. Survey of the field

4.1.1. Semasiology–onomasiology and semantic schematicity

For Cognitive Semantics, one has two possible objects of study. Firstly, one
may ask what forms are available to express a given concept (onomasiol-
ogy or synonymy). Secondly, one can ask what concepts are expressed by a
given form (semasiology or polysemy). Having categorised the research as
onomasiological or semasiological, one may further divide it into the study
of schematic or non-schematic form and meaning. Schematic meanings are
more typically expressed by grammatical forms, such as morphemes and
grammatical constructions while non-schematic meanings are more typi-
cally expressed by lexical forms. The semasiological–onomasiological
division crosses both objects of study. For grammatical meaning, the ono-
masiological study of schematic concepts includes, for example, syntactic
alternations or the choice between grammatical cases. The semasiological
study of grammatical forms looks more closely at a given construction or
morpheme. The same distinction applies to lexical semantics, effectively
drawing a distinction between the study of polysemy networks and near-

synonymy lexical fields. This leads to four research questions – (i) What words do people choose? (ii) What grammatical constructions do people choose? (iii) How do people use a word? (iv) How do people use a grammatical construction? We can label these research questions respectively as lexical synonymy, grammatical synonymy, lexical polysemy, and grammatical polysemy.

Although there is no clear divide between lexical and grammatical forms and obviously even less so between the schematic and non-schematic meaning, the onomasiological–semasiological division is reasonably clear-cut. This 'form-first' or 'concept-first' distinction is arguably fundamental to semantic research (Geeraerts *et al*. 1994; Geeraerts 2006a). However, even here, Glynn (2010) argues that there exists a continuum between these two dimensions, depending on the granularity of the study. This is because all variation in prosody, syntax, and morphology is technically a change in form that can represent a subtle change in meaning. To this extent, the line between semasiology and onomasiology is also not ultimately discernable. Newman (2008b; this vol.) broaches this question in his discussion on the problems of studying the lemma, arguably a very coarse-grained level of analysis. Nevertheless, a methodological division does exist, which demands that research begin with either a concept – examining the various forms that express it, or with a form – examining the various meanings expressed by it.

The four research domains listed above can be used to categorise all research in Cognitive Semantics. The work on culturally determined concepts, semantic frames, cognitive models, and metaphors (Wierzbicka 1985, Kövecses 1986, Lakoff 1987: case study 1) is obviously lexical onomasiology, just as the research in polysemy is typically lexical semasiology (Lakoff 1987: case study 2, Fillmore & Atkins 1992, Cuyckens 1995). The studies in syntactic alternations, popular in Construction Grammar, are clearly grammatical onomasiology. Similarly, Langacker (2000) and Talmy (2000) examine the differences between various grammatical profilings, an inherently schematic-semantic, morpho-syntactic, and onomasiological domain of research. This leaves the semasiological study of individual grammatical forms and syntactic patterns. Such research is less well-known but equally represented. Important studies include Lakoff (1987: case study 3), Janda (1993), Goldberg (1995), Rudzka-Ostyn (1996), Dąbrowska (1997), Geeraerts (1998).

Having established the four research domains of Cognitive Semantics, we need to introduce two methodological variables specific to corpus-

driven analysis - the relative use of observable objective linguistic features and the relative use of statistical techniques.

4.1.2. *Hard statistics–soft statistics and objectivity*

Methodologically, corpus-driven research in semantics is far from being a single established approach. There are different ways of collecting data, analysing the data, and then a wide range of quantitative techniques available for treating the results of these analyses.

First, research in a quantitative semantic study can be restricted to the analysis of formal observable phenomena or it can include semantic non-observable phenomena. This is important since the inclusion of non-observable, and therefore subjectively determined, factors in semantic analysis greatly reduces the objectivity of the study. Moreover, if the analysis is restricted to formal observable characteristics, one has the possibility of the automatic treatment of data. It follows that if the analysis is automatic, one may examine large quantities of examples. As corpora increase in size, and as methods for automatically treating them improve, new possibilities for this line of research are beginning to emerge. Examples of automatic analyses that seek to answer semantic questions include the collostructional analyses of Stefanowitsch & Gries (2003), Gries & Stefanowitsch (2004) and Hilpert (2008) and the lexical onomasiological studies of Arppe (2008) and Levshina *et al*. (forthc.). However, the application of computational techniques, such as word space modelling, to Cognitive Semantic research is also emerging. The Sem•metrix project, led by D. Geeraerts, is an example of such an approach (Peirsman *et al*. 2010).

The lack of semantic richness in formal automatic analysis is offset by the fact that its results can be more reliable, not only due to the objective nature of the analysis, but because a larger sample improves representativity. Nevertheless, despite their merit, such approaches arguably miss much of what is important for semantic research. It is questionable whether linguists will ever be able to sufficiently describe language structure based exclusively upon formal patterns of usage. In light of this, small-scale detailed manual semantic analysis, typical of the usage-feature behavioural profile approach, remains an important line of investigation. Both lines of research, as well as their pros and cons, are treated in this volume.

Determining the degree of statistical sophistication for the treatment of data is not a straightforward question. With small quantities of data, due to

limited resources imposed by transcribing spoken language, scarcity of historical sources, or the practical constraints of fieldwork, one can do little more than count and compare raw numbers. Moreover, the meticulous manual analysis of usage features considerably restricts the number of examples in a dataset, in turn, restricting the efficacy of more advanced statistical techniques. Finally, the technical competence of the author is a real, if not scientific, factor that influences the choice of statistical methods. At least for the current generations of linguists, statistical training is rare, though this is set to change in the near future.

Quantitative approaches to semantic structure within Cognitive Linguistics begin with counting examples of a certain kind and comparing them to examples of another kind (e.g.: Hanegreefs 2004; Davidse *et al*. 2008; Dziwirek & Lewandowska-Tomaszczyk 2009). Such quantitative data may employ tests for statistical significance, such as the t-Test or *Chi*-square test. Other quantitative research employs exploratory techniques such as Cluster Analysis and Correspondence Analysis (Gries 2006a, Divjak 2006, Szelid & Geeraerts 2008, Glynn 2009). These techniques may also be combined with significance testing, but their focus is identifying patterns and associations in the data. Finally, confirmatory statistics such as Linear Discriminant Analysis and Logistic Regression are complex, yet extremely powerful techniques (Gries 2003, Heylen 2005, Speelman and Geeraerts 2010). Confirmatory statistical modelling presents many possibilities for the multivariate description of language, as well as the verification of that description. Glynn & Robinson (in press) offers a survey of the statistical approaches popular in Cognitive Semantics.

4.1.3. Social dimensions of semantics

Special mention must be made of the sociolinguistic element in corpus-driven research. This line of research is of growing importance (Geeraerts 2005, Croft 2009, Geeraerts & Speelman 2010) and is especially significant to quantitative corpus-driven approaches for two reasons. Firstly, sociolinguistics, along with psycholinguistics, has employed some of the most advanced statistics for language study and has, to some extent, paved the way forward for Cognitive Semantics. Secondly, it is precisely this element of language structure that is effectively impossible to account for using introspection and difficult to adequately account for using experimentation. It is for these reasons that the interaction of meaning and society is at the heart

of corpus-driven Cognitive Semantic research. Two recent anthologies, Kristiansen & Dirven (2008), Geeraerts *et al*. (2010), are devoted to the subject.

4.2. Case studies in the field

We can now consider the field of research. Tables 1 and 2 were compiled by collecting the relevant studies from a range of sources. These sources include the anthologies listed in footnote 2, four Cognitive Linguistics journals (*Cognitive Linguistics*, *Annual Review of Cognitive Linguistics*, *Language and Cognition* and *Constructions and Frames*), and three important corpus linguistics journals (*Corpus Linguistics and Linguistic Theory*, *Corpora*, *International Journal of Corpus Linguistics*).[8] Tables 1 and 2 list both the object of study and the method employed. The methodological information offered includes the degree of statistical complexity: counts, collocation (including Collostructional Analysis), exploratory statistics (Hierarchical Cluster Analysis HCA; Principal Component Analysis PCA; Configural Frequency Analysis CFA; Multiple Correspondence Analysis MCA; Multidimensional Scaling MDS), and confirmatory multifactorial analysis (Linear Discriminant Analysis LDA; Profile-Based Analysis PBA; Logistic Regression Analysis LRA; Mixed Effects Logistic Regression MER). These techniques are explained by various authors in Glynn and Robinson (in press).[9] When more than one technique is employed, only the most advanced is listed. The degree of objectivity is also indicated – studies based entirely on formal analysis of observable data versus studies that include at least some non-observable semantic usage-features. The tables also indicate if a study includes extralinguistic dimensions – sociolinguistic, diachronic, or contrastive factors. The studies are sorted chronologically, relative to whether they have a semasiological or onomasiological emphasis and whether they treat lexical or grammatical forms and concepts.

Table 1. Corpus-Driven Cognitive Semantics - Lexical Synonymy and Polysemy

Object		Method		Reference
Lexical Onomasiological				
Verbs of informing (Eng.)	-	counts	sem	Dirven *et al.* 1982
Nouns for clothing (Dch.)	socio.	counts	sem	Geeraerts 1993b; *et al.* 1994
Verbs of starting (Eng.)	-	counts	sem	Schmid 1993
Verbs of response (Eng.)	diach.	counts	sem	Rudzka-Ostyn 1995
Terms for football (Dch.)	socio.	conf. PBA	sem	Geeraerts *et al.* 1999
Abstract nouns (Eng.)	-	counts	sem	Schmid 2000
Verbs of posture (Eng.)	-	counts	form	Newman & Rice 2004a
Concept of ANGER (Eng.)	diach.	counts	sem	Gevaert 2005
Verbs of eating (Eng.)	-	counts	form	Newman & Rice 2006
Verbs of intention (Rus.)	-	explor. HCA	sem	Divjak 2006
Verbs of attempting (Rus.)	-	explor. HCA	sem	Divjak & Gries 2006
Verbs of becoming (Spanish)	-	explor. MDS	sem	Bybee & Eddington 2006
Verbs of separation (Dch., Eng., Germ., Swed.)	contr.	explor. HCA	sem	Majd & Bowerman 2007
Verbs of cognition (Fin.)	-	conf. LRA	form	Arppe 2009
Concepts POSITIVE, NEGATIVE (Hungarian)	socio.	explor. MCA	sem	Szelid & Geeraerts 2008
Verbs of starting (Eng., Rus.)	contr.	explor. HCA	sem	Divjak & Gries 2009
Verbs of reception (Dch.)	diach.	counts	form	Delorge 2009
Adverbs for 'again' (Mnd.)	-	conf. CFA	sem	Jing-Schmidt & Gries 2009
Concepts HAPPY, SAD (Rus.)	-	explor. HCA	sem	Janda & Solovyev 2009; in press
Concepts LOVE, HATE (Pol.)	contr.	colloc. count	form	Dziwirek & Lewandowska 2009
Verbs of posture (Eng.)	diach.	colloc. count	form	Newman 2009
Terms for BOTHER (Eng.)	socio.	conf. LRA	sem	Glynn 2010
Terms for FOOTBALL (Port.)	socio.	conf. PBA	sem	Soares da Silva 2010
Verbs of causation (Dch.)	socio.	conf. LRA	sem	Speelman & Geeraerts 2010
Verbs *come*, *go* (Jap., Kor.)	contr.	counts	sem	Kabata & Lee 2010
Verbs of bothering (Eng.)	socio.	conf. MER	sem	Glynn in press a
Verbs of cognition (Pol.)	socio.	conf. LRA	sem	Fabiszak *et al.* in press
Verbs of causation (Dch.)	socio.	conf. MER	form	Levshina *et al.* in press
Verbs of possibility (Eng., Fr.)	contr.	conf. LRA	sem	Deshors & Gries in press
Concept LIBERTY (Eng.)	socio.	conf. LRA	sem	Glynn forthc.
Lexical Semasiological				
Verb *ask* (Eng.)	diach.	counts	sem	Rudzka-Ostyn 1989
Verb *need to* (Eng.)	socio.	colloc. count	sem	Nokkonen 2006
Verb *run* (Eng.)	-	explor. HCA	sem	Gries 2006a
Verb, Noun *hassle* (Eng.)	socio.	explor. MCA	sem	Glynn 2009
Verb, Noun *bother* (Eng.)	socio.	conf. LRA	sem	Glynn this volume
Adjective *deep* (Germ.)	-	colloc.	sem	Zeschel this volume
Verb *annoy* (Eng.)	socio.	explor. MCA	sem	Glynn submitted

Legend: Adj. - Adjective; Adv. - Adverb; Alt. - Alternation; Cx. - Construction; Imperf. - Imperfective; Perf. - Perfective; Dch. - Dutch; Eng. - English; Fin. - Finnish; Fr. - French; Germ. - German; Jap. - Japanese; Kor. - Korean; Mnd. - Mandarin; Pol. - Polish; Port. - Portugese; Rus. - Russian; Swed. - Swedish; Ukr. - Ukrainian; contr. - contrastive; diach. - diachronic; socio. - sociolinguistic; confirm. - confirmatory statistics; explor. - exploratory statistics; sem - semantic annotation, form - formal annotation.

Table 2. Corpus-Driven Cognitive Semantics – Grammatical Synonymy and Polysemy

Object		Method		Reference
Grammatical Onomasiological				
Alt. Phrasal Verb syntax (Eng.)	-	counts	form	Gries 1999
Alt. Future Cxs (Eng.)	-	counts	form	Szmrecsanyi 2003
Alt. Adj. Syntax (Eng.)	-	confirm. LDA	sem	Wulff 2003
Alt. Adj. Inflection (Eng.)	-	explor. PCA	form	Gries 2003
Aspect Prepositions (Eng.)	-	counts	form	Newman & Rice 2004b
Alt. Adj. Inflection (Dch.)	socio.	confirm. LRA	form	Tummers *et al*. 2005
Alt. Middle Field Syntax (Germ.)	socio.	confirm. LRA	sem	Heylen 2005
Alt. *go* V–*go and* V Cxs (Eng.)	-	colloc. collostr.	form	Wulff 2006
Grammatical Cases in Slavic	contr.	explor. MDA	sem	Clancy 2006
Alt. Dative	socio.	confirm. LRA	sem	Bresnan *et al*. 2007
Alt. Presentative Cxs (Dch.)	socio.	confirm. LRA	sem	Grondelaers *et al*. 2007; 2008
Temp. Adv. Clause Syntax (Eng.)	-	confirm. LRA	sem	Diessel 2008
Alt. Nominal – Clausal Compls (Eng.)	-	colloc. collostr.	form	Wiechmann 2008b
Alt. Act-Pass. Voice; Alt. Future Cxs; Alt. Phrasal Verb syntax (Eng.)	socio.	confirm. CFA	form	Stefanowitsch & Gries 2008
Alt. Dep. Clause syntax (Dch.)	socio.	confirm. LRA	sem	De Sutter 2009; *et al.* 2008
Alt. Adj. Inflection (Dch.)	socio.	colloc. collostr.	form	Speelman *et al*. 2009
Alt. Genitive Cxs (Eng.)	socio.	confirm. LRA	sem	Szmrecsanyi 2010
Alt. Imperf.–Perf. Cxs (Rus.)	-	confirm. MER	form	Divjak this vol.
Alt. Case – Adposition (Estonian)	socio.	confirm. LRA	sem	Klavan in press
Grammatical Semasiological				
Case Dative (Pol.)	-	counts	sem	Rudzka-Ostyn 1996
Cxs. *think nothing of* Gerund; Imper.; Progr.; Ditrans. (Eng.)	-	colloc. collostr.	form	Stefanowitsch & Gries 2003
Cx. *into*-Causative (Eng.)	-	confirm. CFA	form	Gries & Stefanowitsch 2004
Cx. *into*-Causative (Eng.)	socio.	colloc. collostr.	form	Wulff *et al*. 2007
Cxs Future (Eng., Dan., Swed., Germ.)	diach. contr.	colloc. collostr.	form	Hilpert 2008a
Cx. Comparative (Eng.)	-	colloc. collostr.	sem	Hilpert 2008b
Aux. *shall*; Aspect Perfect (Eng.)	diach.	explor. HCA	form	Gries & Hilpert 2008
Cx. *good* PrP (Eng.)	-	counts	sem	Zeschel 2009
Cx. *mit*-Predicative (Germ.)	-	confirm. CFA	form	Hilpert 2009
Cx. Dative (Dch.)	diach.	colloc. collostr.	form	Colleman 2009, this vol.
Cx. Raising (Eng., Dch.)	diach. contr.	colloc. collostr.	form	Noël & Colleman 2010
Cx. Verb Poss. *way* (Eng.)	-	explor. HCA	form	Gries & Stefanowitsch 2010
Cx. Benefactive Ditransitive (Eng.)	socio.	counts	form	Colleman 2010
Cx. V – NP (Eng.)	-	colloc. collostr.	form	Wulff this vol.
Cx. Gerund Compl. Clause (Eng.)	-	colloc. collostr.	form	Hilpert this vol.
Cx. *for* Durative Adv. (Eng.)	-	colloc. collostr.	form	Fuhs this vol.
Suffix dissposseive –*ont* (Dch.)	diach	counts	sem	Delorge & Colleman in press
Cx. Epistemic stance (Eng.)	socio	confirm. MER	sem	Glynn & Krawczak submitted

The first thing one will notice looking at the tables is that for each object of study, there is a tendency towards more sophisticated quantitative techniques over time. Just as noticeable is the tendency for semantic analysis in lexical research contrasted by formal analysis in grammatical research. Although this is in part due to a large amount of collocation studies of grammatical constructions, there is no reason to suppose that collocation should not be equally applied to the study of lexical semantics. As Zeschel (this vol.) shows, the collocation tradition, which was originally lexically orientated, still has a great deal to offer the study of lexical semantics. Noteworthy as well is a small increase in the amount of studies including sociolinguistic, diachronic, and contrastive parameters in their research.

The tables also reveal the biases in research interests across the research paradigm. In lexical research, verbs obviously have an over representation, nouns, adjectives, and adverbs being poorly represented. Also, and somewhat surprisingly, seeing the history of Cognitive Semantics, prepositions are entirely absent. Some of this bias is because verbs, due to their argument structure and its participants, offer more observable features for analysis, facilitating quantitative study. Overcoming the difficulties of studying other parts of speech should, nevertheless, be a goal in the research field. Equally noticeable is the emphasis upon onomasiological research. Again, there is a likely methodological reason. It is relatively easy to distinguish the use of forms compared to distinguishing uses, especially since formal and extralinguistic factors may often contribute a lot to explaining onomasiological variation. The research field needs to approach the difficulties of semasiological lexical description and this volume offers two contributions, Zeschel (this vol.) and Glynn (this vol.), that attempt this.

The grammatical research also shows clear biases. Obviously, the popularity of the Collostructional Analysis accounts for much of this. However, the fact that effectively all onomasiological research examines syntactic alternations and all semasiological research looks at syntactic-lexical pairing is a clear and serious shortcoming. Although there are some important exceptions to this trend, such as the work on adjectival inflections and aspectual categories, the field must move towards other kinds of grammatical semantics. The semantics encoded by phonological structures, such a prosody, are entirely absent. The lack of research in grammatical case is especially surprising, seeing the rich tradition of this in Cognitive Linguistics. This volume corrects this imbalance with Divjak's (this vol.) research on aspect. However, both the semasiological and onomasiological research in grammatical semantics needs to broaden its field of research.

Perhaps, the most striking gap in the research is the lack of work on discourse questions and broad culturally-determined concepts, such as those proposed by Lakoff (1987). The first steps have been taken in the study of Idealised Cognitive Models (Dziwirek & Lewandowska 2009, Glynn forthc.), but this volume does not offer any such research. Extending this method to more culturally-orientated concepts is a clear goal for future research. The study of discourse is difficult due to the lack of adequate spoken corpora. Fischer (this vol.) offers an important contribution on this front by developing a corpus especially designed to capture such discourse structures. As the field matures, such lines of inquiry will certainly attract more attention. The next step is to develop the methods to answer broader semantic research questions. This volume makes a small, but important step, in this direction.

Notes

1. I would like to thank G. Gilquin, M. Hilpert, and C. Paradis for their help. All shortcomings remain my own.
2. Testimony to the surge in corpus-driven research is the large number of anthologies, devoted entirely or substantially, to corpus-based Cognitive Linguistics. A list of these includes Gries & Stefanowitsch (2006), Stefanowitsch & Gries (2006), Gonzalez-Marquez *et al*. (2007), Kristiansen & Dirven (2008), Zeschel (2008), Lewandowska-Tomaszczyk & Dziwirek (2009), Gilquin (2009), Gries *et al*. (2009). Geeraerts *et al*. (2010), Rice & Newman (2010), Marzo *et al*. (2010), Glynn & Robinson (in press), and Gries & Divjak (forthc.). Geeraerts (2006b) offers a detailed description of the generalisation of this trend across the field.
3. R. Gibbs and M. Tomasello lead the field. One cannot even begin to list the most important contributions here, but early studies include Gibbs (1990) and Tomasello (2000).
4. Although one may interpret Langacker's position in this way, from a strictly cognitive perspective, the idea of competence is not at all relevant. Cf. Paradis (2003) for a discussion on the notion of competence.
5. The use of frequency as a means for operationalising grammar deserves discussion beyond the scope of this introduction. Stefanowitsch (this vol.) shows its importance at a methodological level and Bybee (2007), amongst others, has fully integrated this notion into her research.

6. Langacker (1987: 100) is explicit about the relationship between entrench-ment and frequency. Schmid (2000: 39) is still more explicit, proposing, what he terms, the Corpus-to-Cognition Principle – "Frequency in text instantiates entrenchment in the cognitive system." Schmid has since, however, distanced himself from this position.

7. Gries *et al*. (2005, 2010), Arppe & Järvikivi (2007) and Wulff (2009b) exam-ine this question in empirical terms. Jones *et al*. (2007) also offer an interest-ing means of operationalising these kinds of issues. Geeraerts (2000) directly addresses the issue of salience in lexical semantics.

8. For sake of brevity, recent monographs in the field have not been included. For the reader's reference, books devoted to the field include Geeraerts *et al*. (1999), Schmid (2000), Fischer (2000), Gries (2003), Mukherjee (2005), Szmrecsanyi (2006), Hilpert (2008), Arppe (2008), Wulff (2009a), Divjak (2010), and Gilquin (2010).

9. Specific references for these techniques include: Collostructional Analysis - Hilpert (in press), Stefanowitsch & Gries (2003), Gries & Stefanowitsch (2004), Hierarchical Cluster Analysis - Divjak & Fieller (in press), Logistic Regression Analysis - Baayen (2008) and Speelman (in press), Mixed Effects Regression - Baayen (in press), Multiple Correspondence Analysis - Glynn (in press b).

References

Arppe, Antti
   2008   *Univariate, Bivariate and Multivariate Methods in Corpus-Based Lexicography – A study of synonymy*. Helsinki: University of Hel-sinki.
   2009   Linguistic choices vs. probabilities - how much and what can lin-guistic theory explain? In *The Fruits of Empirical Linguistics.* Vol. 1. *Process*, Sam Featherston and Susanne Winkler (eds.), 1-14. Ber-lin / New York: Mouton de Gruyter.
Arppe, Antti and Juhani Järvikivi
   2007   Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Lin-guistic Theory* 3: 131–159.
Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert, Arne Zeschel
   2010   Cognitive corpus linguistics: Five points of debate on current theory and methodology. *Corpora* 5: 1-27.

Baayen, R. Harald
  2008   *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
  in press Multinomial mixed models. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.
Bresnan, Joan, Anna Cueni, Tatiana Nikitina and Harald Baayen
  2007   Predicting the Dative alternation. In *Cognitive Foundations of Interpretation*, Gosse Bouma, Irene Kraemer and Joost Zwarts (eds.), 69–94. Amsterdam: Royal Netherlands Academy of Science.
Bybee, Joan
  2007   *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.
Bybee, Joan and David Eddington
  2006   A usage-based approach to Spanish verbs of becoming. *Language* 82: 323-354.
Clancy, Steven J.
  2006.   The topology of Slavic case: Semantic maps and multidimensional scaling. *Glossos* 7: 1-28.
Colleman, Timothy
  2009   The semantic range of the Dutch double object construction: A collostructional perspective. *Constructions and Frames* 1: 190-221.
  2010   Lectal variation in constructional semantics: "Benefactive" ditransitives in Dutch. In *Advances in Cognitive Sociolinguistics*, Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman (eds.), 191-221. Berlin / New York: Mouton de Gruyter.
Croft, William
  2009   Toward a social Cognitive Linguistics. In *New Directions in Cognitive Linguistics*, Vyvyan Evans and Stéphanie Pourcel (eds.), 395-420. Amsterdam / Philadelphia: John Benjamins.
Cuyckens, Hubert
  1995   Family resemblance in the Dutch spatial prepositions *door* and *langs*. *Cognitive Linguistics* 6: 183-207.
Dąbrowska, Ewa
  1997   *Cognitive Semantics and the Polish Dative*. Berlin / New York: Mouton de Gruyter.
Davidse, Kristin, Lieselotte Brems and Liesbeth De Smedt
  2008   Type noun uses in the English NP: A case of right to left layering. *International Journal of Corpus Linguistics* 13: 139–168.

De Sutter, Gert
   2009   Towards a multivariate model of grammar: The case of word order
             variation in Dutch clause final verb clusters. In *Describing and
             Modeling Variation in Grammar*, Andreas Dufter, Jürg Fleischer
             and Guido Seiler (eds.), 225-254. Berlin / New York: Mouton de
             Gruyter.
De Sutter, Gert, Dirk Speelman and Dirk Geeraerts
   2008   Prosodic and syntactic-pragmatic mechanisms of grammatical vari-
             ation: The impact of a postverbal constituent on the word order in
             Dutch clause final verb clusters. *International Journal of Corpus
             Linguistics* 13: 194-224.
Deignan, Alice
   2005   *Metaphor and Corpus Linguistics*. Amsterdam / Philadelphia: John
             Benjamins.
   2009   Searching for metaphorical patterns in corpora. In *Contemporary
             Corpus Linguistics*, Paul Baker (ed.), 9-31. London: Continuum.
Delorge, Martine
   2009   A diachronic corpus study of the constructional behaviours of recep-
             tion verbs in Dutch. In *Studies in Cognitive Corpus Linguistics*, Bar-
             bara Lewandowska-Tomaszczyk and Katarzyna Dziwirek (eds.),
             249 -272. Frankfurt/Main: Peter Lang.
Delorge, Martine and Timothy Colleman
   in press   The polysemy of Dutch 'transfer' constructions: The case of ont-
             verbs. In *Polysemy and Synonymy. Corpus methods and applications
             in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.).
             Amsterdam / Philadelphia: John Benjamins.
Deshors, Sandra and Stefan Th. Gries
   in press   A case for the multifactorial assessment of learner language: The
             uses of *may* and *can* in French-English interlanguage. In *Polysemy
             and Synonymy. Corpus methods and applications in Cognitive Lin-
             guistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam /
             Philadelphia: John Benjamins.
Dirven, René and John Taylor
   1988   The conceptualisation of vertical space in English: The case of
             "tall". In *Topics in Cognitive Linguistics*, Brygida Rudzka-Ostyn
             (ed.), 379-402. Amsterdam / Philadelphia: John Benjamins.
Dirven, René, Louis Goossens, Yvan Putseys and Emma Vorlat
   1982   *The Scene of Linguistic Action and its Perspectivization by* SPEAK,
             TALK, SAY, and TELL. Amsterdam / Philadelphia: John Benjamins.

Diessel, Holger
    2008   Iconicity of sequence. A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics* 19: 457–482.

Divjak, Dagmar
    2006   Ways of intending: A corpus-based cognitive linguistic approach to near-synonyms in Russian. In *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*, Stefan Th. Gries and Anatol Stefanowitsch (eds.), 19–56. Berlin / New York: Mouton de Gruyter.
    2010   *Structuring the Lexicon: A Clustered Model for Near-synonymy*. Berlin / New York: Mouton de Gruyter.

Divjak, Dagmar and Nick Fieller
    in press   Clustering linguistic data. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.

Divjak, Dagmar and Stefan Th. Gries
    2006   Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2: 23–60.
    2009   Corpus-based Cognitive Semantics: A contrastive study of phasal verbs in English and Russian. In *Studies in Cognitive Corpus Linguistics*, Barbara Lewandowska-Tomaszczyk and Katarzyna Dziwirek (eds.), 273-296. Frankfurt/Main: Peter Lang.

Durkin, Kevin and Jocelyn Manning
    1989   Polysemy and the subjective lexicon: Semantic relatedness and the salience of intraword senses. *Journal of Psycholinguistic Research* 18: 577-612.

Dziwirek, Katarzyna and Barbara Lewandowska-Tomaszczyk
    2009   Love and hate. Unique transitive emotions in English and Polish. In *Studies in Cognitive Corpus Linguistics*, Barbara Lewandowska-Tomaszczyk and Katarzyna Dziwirek (eds.), 297-318. Frankfurt/Main: Peter Lang.

Fabiszak, Małgorzata, Anna Hebda, Iwona Kokorniak and Karolina Krawczak
    in press   The prefix semantics and grammar in the Polish verb 'to think'. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.

Fillmore, Charles
    1985   Frames and the Semantics of Understanding. *Quaderni di Semantica* 6: 222-254.

Fillmore, Charles and Beryl Atkins

    1992   Toward a frame-based lexicon: The semantics of RISK and its neighbours. In *Frames, Fields, and Contrasts: New essays in semantic and lexical organization*, Adrienne Lehrer and Eva Kittay (eds.), 75-102. London: Lawrence Erlbaum Associates.

Fischer, Kerstin

    2000   *From Cognitive Semantics to Lexical Pragmatics: The functional polysemy of discourse particles*. Berlin / New York: Mouton de Gruyter.

Geeraerts, Dirk

    1993a   Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics* 4: 223-272.

    1993b   Generalised onomasiological salience. In *Perspectives on Language and Conceptualization* (Special ed. *Belgian Journal of Linguistics* 8), Jan Nuyts and Eric Pederson (eds.), 43-56. Brussels: Presses Universitaires de Bruxelles.

    1998   The semantic structure of the indirect object in Dutch. In *The Dative.* Vol. 2. *Theoretical and contrastive studies*, Willy van Langendonck and William van Belle (eds.), 185-210. Amsterdam / Philadelphia: John Benjamins.

    2000   Salience phenomena in the lexicon. A typology. In *Meaning and Cognition*, Liliana Albertazzi (ed.), 125-136. Amsterdam / Philadelphia: John Benjamins.

    2005   Lectal data and empirical variation in Cognitive Linguistics. In *Cognitive Linguistics. Internal dynamics and interdisciplinary interactions*, Francisco José Ruiz de Mendoza Ibáñez and Sandra Peña Cervel (eds.), 163-189. Berlin: Mouton de Gruyter.

    2006a   *Words and Other Wonders: Papers on lexical and semantic topics*. Berlin / New York: Mouton de Gruyter.

    2006b   Methodology in Cognitive Linguistics. In *Cognitive Linguistics: Current applications and future perspectives*, Gitte Kristiansen, Michel Achard, René Dirven and Francisco J. Ruiz de Mendoza Ibañez (eds.), 21-50. Berlin / New York: Mouton de Gruyter.

    2007   Interview: Cognitive sociolinguistics and the sociology of Cognitive Linguistics, by Juana Marín-Arrese, *Annual Review of Cognitive Linguistics* 5: 289–305.

    2010   Recontextualizing grammar: Underlying trends in thirty years of Cognitive Linguistics. In *Cognitive Linguistics in Action: From theory to application and back*, Elżbieta Tabakowska, Micha Choinski and Łukasz Wiraszka (eds.), 71–102. Berlin / New York: Mouton de Gruyter.

Geeraerts, Dirk and Speelman, Dirk
2010   Heterodox concept features and onomasiological heterogeneity in dialects. In *Advances in Cognitive Sociolinguistics*, Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman (eds.), 21-40. Berlin / New York: Mouton de Gruyter.

Geeraerts, Dirk, Gitte Kristiansen and Yves Peirsman (eds.)
2010   *Advances in Cognitive Sociolinguistics*. Berlin / New York: Mouton de Gruyter.

Geeraerts, Dirk, Stefan Grondelaers and Dirk Speelman
1999   *Convergentie en Divergentie in de Nederlandse Woordenschat.* Amsterdam: Meertens Instituut.

Geeraerts, Dirk, Stefan Grondelaers and Peter Bakema
1994   *The Structure of Lexical Variation. Meaning, naming, and context*. Berlin / New York: Mouton de Gruyter.

Gevaert, Caroline
2005   The anger is heat question: Detecting cultural influence on the conceptualization of anger through diachronic corpus analysis. In *Perspectives on Variation. Sociolinguistic, historical, comparative*, Nicole Delbecque, Johan van der Auwera and Dirk Geeraerts (eds.), 195-208. Berlin / New York: Mouton de Gruyter.

Gibbs, Raymond
1990   Psycholinguistic studies on the conceptual basis of idiomaticity. *Cognitive Linguistics* 1: 417-451.

Gilquin, Gaëtanelle
2006   The place of prototypicality in corpus linguistics. Causation in the hot seat. In *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*, Stefan Th. Gries and Anatol Stefanowitsch (eds.), 159-191. Berlin / New York: Mouton de Gruyter.
2010   *Corpus, Cognition and Causative Constructions*. Amsterdam / Philadelphia: John Benjamins.

Gilquin, Gaëtanelle (ed.)
2009   *Corpora and Experimental Methods* (Special ed. *Corpus Linguistics and Linguistic Theory* 5). Berlin / New York: Mouton de Gruyter.

Gilquin, Gaëtanelle and Stefan Th. Gries
2009   Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5: 1-26.

Givón, Talmy
1991   Markedness in grammar: distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language* 15: 335-370.

Glynn, Dylan
 2009 Polysemy, syntax, and variation. A usage-based method for Cognitive Semantics. In *New Directions in Cognitive Linguistics*, Vyvyan Evans and Stéphanie Pourcel (eds.), 77-106. Amsterdam / Philadelphia: John Benjamins.
 2010 Synonymy, lexical fields, and grammatical constructions. A study in usage-based Cognitive Semantics. In *Cognitive Foundations of Linguistic Usage-Patterns*, Hans-Jörg Schmid and Susanne Handl (eds.), 89-118. Berlin / New York: Mouton de Gruyter.
 in press a Lexical synonymy and onomasiological structure in a usage-based model of language. A multifactorial case study. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.
 in press b Multiple correspondence analysis. Exploring correlations in multifactorial data. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.
 submitted Semantics of sociolinguistic variation. A quantitative study of dialect effects on the polysemy of *annoy*. *Annual Review of Cognitive Linguistics*.
 forthc. Usage-based cognitive models. Lexical synonymy, conceptual structure and American LIBERTY. *Cognitive Linguistics*.
Glynn, Dylan and Justyna Robinson (eds.)
 in press *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics.* Amsterdam / Philadelphia: John Benjamins.
Glynn, Dylan and Karolina Krawczak
 submitted Social cognition, Cognitive Grammar and corpora. A multifactorial approach to subjectification. *Language and Cognition*
Goldberg, Adele
 1995 *Constructions. A Construction Grammar approach to argument structure*. London: University of Chicago Press.
Gonzalez-Marquez, Monica, Irene Mittleberg, Seanna Coulson and Michael Spivey (eds.)
 2007 *Methods in Cognitive Linguistics*. Amsterdam / Philadelphia: John Benjamins.
Goossens, Louis
 1990 Metaphtonymy: The interaction of metaphor and metonymy in figurative expressions for linguistic action. *Cognitive Linguistics* 1: 323-34.

Gries, Stefan Th.

    1999   Particle movement: A cognitive and functional approach. *Cognitive Linguistics* 10: 105-145.

    2001   A multifactorial analysis of syntactic variation: Particle movement revisited. *Journal of Quantitative Linguistics* 8: 33-50.

    2003   *Multifactorial Analysis in Corpus Linguistics: A study of particle placement.* London: Continuum Press.

    2006a   Corpus-based methods and Cognitive Semantics: The many senses of *to run*. In *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*, Stefan Th. Gries and Anatol Stefanowitsch (eds.), 57–99. Berlin / New York: Mouton de Gruyter.

    2006b   Exploring variability within and between corpora: Some methodological considerations. *Corpora* 1: 109–51.

    2008   Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13: 403-437.

Gries, Stefan Th. and Anatol Stefanowitsch (eds.)

    2006   *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis.* Berlin / New York: Mouton de Gruyter.

Gries, Stefan Th. and Anatol Stefanowitsch

    2004   Extending collostructional analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9: 97-129.

    2010   Cluster analysis and the identification of collexeme classes. In *Empirical and Experimental Methods in Cognitive/Functional Research*, Sally Rice and John Newman (eds.), 73-90. Stanford, CA: CSLI.

Gries, Stefan Th. and Dagmar Divjak (eds.)

    forthc.   *Frequency Effects in Language*. Berlin / New York: Mouton de Gruyter.

Gries, Stefan Th. and Martin Hilpert

    2008   The identification of stages in diachronic data: Variability-based neighbor clustering. *Corpora* 3: 59-81.

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: 635-676.

    2010   Converging evidence II: More on the association of verbs and constructions. In *Empirical and Experimental Methods in Cognitive/Functional Research*, Sally Rice and John Newman (eds.), 59-72. Stanford: CSLI.

Gries, Stefan Th., Stefanie Wulff and Mark Davies (eds.)
    2009   *Corpus Linguistic Applications: Current studies, new directions*. Amsterdam: Rodopi.

Grondelaers, Stefan, Dirk Geeraerts and Dirk Speelman
    2007   A case for a cognitive corpus Linguistics. In *Methods in Cognitive Linguistics*, Monica Gonzalez-Marquez, Irene Mittleberg, Seanna Coulson and Michael Spivey (eds.), 149-169. Amsterdam / Philadelphia: John Benjamins.

Grondelaers, Stefan, Dirk Speelman and Dirk Geeraerts
    2008   National variation in the use of *er* "there". Regional and diachronic constraints on cognitive explanations. In *Cognitive Sociolinguistics: Language variation, cultural models, social systems*, Gitte Kristiansen and René Dirven (eds.), 153-204. Berlin / New York: Mouton de Gruyter.

Hanegreefs, Hilde
    2004   The Dutch change-of-state copula 'worden' and its Spanish counterparts: A matter of aspect and voice. *Annual Review of Cognitive Linguistics* 1: 29.

Heylen, Kris
    2005   A quantitative corpus study of German word order variation. In *Linguistic Evidence: Empirical, theoretical and computational perspectives*, Stephan Kepser and Marga Reis (eds.), 241-264. Berlin / New York: Mouton de Gruyter.

Heylen, Kris, José Tummers and Dirk Geeraerts
    2008   Methodological issues in corpus-based Cognitive Linguistics. In *Cognitive Sociolinguistics: Language variation, cultural models, social systems*, Gitte Kristiansen and René Dirven (eds.), 91-128. Berlin / New York: Mouton de Gruyter.

Hilpert, Martin
    2008a   *Germanic Future Constructions. A usage-based approach to language change*. Amsterdam / Philadelphia: John Benjamins.
    2008b   The English comparative – language structure and language use. *English Language and Linguistics* 12: 395–417.
    2009   The German *mit*-predicative construction. *Constructions and Frames* 1: 29-55.
    in press   Collostructional Analysis: Measuring associations between constructions and lexical elements. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.

Janda, Laura
    1993   *A Geography of Case Semantics: The Czech dative and the Russian instrumental*. Berlin / New York: Mouton de Gruyter.

Janda, Laura and Valery Solovyev
    2009   What constructional profiles reveal about synonymy: A case study of the Russian words for SADNESS and HAPPINESS. *Cognitive Linguistics* 20: 367-393.
    in press   Antonyms as types of synonyms: Russian emotion terms. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.

Jing-Schmidt, Zhuo and Stefan Th. Gries
    2009   Schematic meaning and pragmatic inference: The Mandarin adverbs *hai*, *you*, and *zai*. *Corpora* 4: 33-70.

Jones, Steven, Carita Paradis, Lynne Murphy and Caroline Willners
    2007   Googling for 'opposites': A web-based study of antonym canonicity. *Corpora* 2: 129-155.

Kabata, Kaori and Jeong-Hwa Lee
    2010   'Come' and 'go' serial verbs in Japanese and Korean: A corpus-based comparative study. In *Empirical and Experimental Methods in Cognitive/Functional Research*, Sally Rice and John Newman (eds.), 109 -126. Stanford: CSLI.

Klavan, Jane
    in press   A multifactorial analysis of grammatical synonymy: The case of the Estonian adessive and the adposition *peal* 'on'. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.

Kövecses, Zoltán
    1986   *Metaphors of Anger, Pride, and Love. A lexical approach to the structure of concepts*. Amsterdam / Philadelphia: John Benjamins.

Kristiansen, Gitte and René Dirven (eds.)
    2008   *Cognitive Sociolinguistics: Language variation, cultural models, social systems*. Berlin / New York: Mouton de Gruyter.

Lakoff, George
    1982   Categories. An essay in Cognitive Linguistics. In *Linguistics in the Morning Calm*, Linguistic Society of Korea (ed.), 139-194. Seoul: Hanshin.
    1987   *Women, Fire, and Dangerous Things. What categories reveal about the mind.* London: University of Chicago Press.

Langacker, Ronald
   1987   *Foundations of Cognitive Grammar*. Vol. 1. *Theoretical prerequi-sites*. Stanford: Stanford University Press.
   1988   A usage-based model. In *Topics in Cognitive Linguistics*, Brygida Rudzka-Ostyn (ed.), 127-161. Amsterdam / Philadelphia: John Benjamins.
   2000   *Grammar and Conceptualization*. Berlin / New York: Mouton de Gruyter.
Lehrer, Adrienne
   1982/2000 *Wine and Conversation*. Bloomington: Indiana University Press/ Oxford: Oxford University Press.
Levshina, Natalia, Dirk Speelman and Dirk Geeraerts
   in press Dutch causative constructions: Quantification of meaning and mean-ing of quantification. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.
Lewandowska-Tomaszczyk, Barbara and Katarzyna Dziwirek (eds.)
   2009   *Studies in Cognitive Corpus Linguistics*. Frankfurt/Main: Lang.
Majid, Asifa and Melissa Bowerman
   2007   Cutting and breaking events: A cross-linguistic perspective. *Cogni-tive Linguistics* 18: 133-152.
Marzo, Stefania, Kris Heylen and Gert De Sutter (eds.)
   2010   *Corpus Studies in Contrastive Linguistics* (Special ed. *International Journal of Corpus Linguistics* 15). Amsterdam / Philadelphia: John Benjamins.
Mukherjee, Joybrato
   2005   *English Ditransitive Verbs: Aspects of theory, description and a usage-based model*. Amsterdam / New York: Rodopi.
Newman, John
   2008a   Spoken corpora: Rationale and application. *Taiwan Journal of Lin-guistics* 6: 27-58.
   2008b   Aiming low in linguistics: Low-level generalizations in corpus-based research. *11th International Symposium on Chinese Lan-guages and Linguistics*, Chiao Tung University, Taiwan.
   2009   English posture verbs: An experientially grounded approach. *Annual Review of Cognitive Linguistics* 7: 30–57.
Newman, John and Sally Rice
   2004a   Aspect in the making: A corpus analysis of English aspect-marking prepositions. In *Language, Culture and Mind*, Suzanne Kemmer and Michel Achard (eds.), 313-327. Stanford: CSLI.

2004b   Patterns of usage for English SIT, STAND, and LIE: A cognitively-inspired exploration in corpus linguistics. *Cognitive Linguistics* 15: 351-396.

2006   Transitivity schemas of English EAT and DRINK in the BNC. In *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*, Stefan Th. Gries and Anatol Stefanowitsch (eds.), 225-260. Berlin / New York: Mouton de Gruyter.

Noël, Dirk and Timothy Colleman

2010   *Believe*-type raising-to-object and raising-to-subject verbs in English and Dutch: A contrastive investigation in diachronic Construction Grammar. *Corpus Studies in Contrastive Linguistics*, Stefania Marzo, Kris Heylen and Gert De Sutter (eds.), 157–182. Amsterdam / Philadelphia: John Benjamins.

Nokkonen, Soili

2006   The semantic variation of NEED TO in four recent British English corpora. *International Journal of Corpus Linguistics* 11: 29-71.

Paradis, Carita

2003   Is the notion of *linguistic competence* relevant in Cognitive Linguistics? *Annual review of Cognitive Linguistics* 1: 247-271.

Peirsman, Yves, Kris Heylen and Dirk Geeraerts

2010   Applying word space models to sociolinguistics. In *Advances in Cognitive Sociolinguistics*, Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman (eds.), 111-138. Berlin / New York: Mouton de Gruyter.

Rice, Sally and John Newman (eds.)

2010   *Empirical and Experimental Methods in Cognitive/Functional Research*. Stanford: CSLI.

Rudzka-Ostyn, Brygida

1989   Prototypes, schemas, and cross-category correspondences: The case of ask. In *Prospects and Problems of Prototype Theory* (Special ed. *Linguistics* 27), Dirk Geeraerts (ed.), 613-661. Berlin / New York: Mouton de Gruyter.

1995   Metaphor, schema, invariance: The case of verbs of answering. In *By Word of Mouth. Metaphor, Metonymy, and Linguistic Action from a Cognitive Perspective*, Louis Goossens, Paul Pauwels, Brygida Rudzka-Ostyn, Anne-Marie Simon-Vandenbergen and Johan Vanparys (eds.), 205-244. Amsterdam / Philadelphia: John Benjamins.

1996   The Polish Dative. In *The Dative*. Vol. 1. *Descriptive studies*. William van Belle and Willy van Langendonck (eds.), 341-394. Amsterdam / Philadelphia: John Benjamins.

Sandra, Dominiek and Sally Rice
    1995   Network analysis of prepositional meaning: mirroring whose mind - the linguist's or the language user's? *Cognitive Linguistics* 6: 89-130.

Schmid, Hans-Jörg
    1993   *Cottage and Co., Idea, Start vs. Begin. Die Kategorisierung als Grundprinzip einer differenzierten Bedeutungsbeschreibung*. Tübingen: Niemeyer.
    2000   *English Abstract Nouns as Conceptual Shells. From corpus to cognition*. Berlin / New York: Mouton de Gruyter.

Schulze, Rainer
    1988   A short story of *down*. In *Understanding the Lexicon. Meaning, sense, and world knowledge in lexical semantics*, Werner Hüllen and Rainer Schulze (eds.), 395-414. Tübingen: Niemeyer.

Soares da Silva, Augusto
    2010   Measuring and parameterizing lexical convergence and divergence between European and Brazilian Portuguese. In *Advances in Cognitive Sociolinguistics*, Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman (eds.), 41-84. Berlin / New York: Mouton de Gruyter.

Speelman, Dirk
    in press   Modeling near-synonymy with logistic regression. In *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*, Dylan Glynn and Justyna Robinson (eds.). Amsterdam / Philadelphia: John Benjamins.

Speelman, Dirk and Dirk Geeraerts
    2010   Causes for causatives: the case of Dutch 'doen' and 'laten'. In *Causal Categories in Discourse and Cognition*, Ted Sanders and Eve Sweetser (eds.), 173-204. Berlin / New York: Mouton de Gruyter.

Speelman, Dirk, José Tummers and Dirk Geeraerts
    2009   Lexical patterning in a Construction Grammar: The effect of lexical co-occurrence patterns on the inflectional variation in Dutch attributive adjectives. *Constructions and Frames* 1: 87–118.

Speelman, Dirk, Stefan Grondelaers and Dirk Geeraerts
    2003   Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities* 37: 317-337.

Stefanowitsch, Anatol
    2006   Negative evidence and the raw frequency fallacy. *Corpus Linguistics and Linguistic Theory* 2: 61-77.
    2008   Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19: 513-531.

Stefanowitsch, Anatol and Stefan Th. Gries
    2003   Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8: 209-243.
    2008   Register and constructional semantics: A collostructional case study. In *Cognitive Sociolinguistics*, Gitte Kristiansen and René Dirven (eds.), 129–152. Berlin / New York: Mouton de Gruyter.
Stefanowitsch, Anatol and Stefan Th. Gries (eds.)
    2006   *Corpus-Based Approaches to Metaphor and Metonymy*. Berlin / New York: Mouton de Gruyter.
Szelid, Veronika and Dirk Geeraerts
    2008   Usage-based dialectology. Emotion concepts in the Southern Csango dialect. *Annual Review of Cognitive Linguistics* 6: 23-49.
Szmrecsanyi, Benedikt
    2003   *Be going to* versus *will/shall*: Does syntax matter? *Journal of English Linguistics* 31: 295-323.
    2006   *Morphosyntactic Persistence in Spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin / New York: Mouton de Gruyter.
    2010   The English genitive alternation in a Cognitive Sociolinguistics perspective. In *Advances in Cognitive Sociolinguistics*, Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman (eds.), 141-166. Berlin / New York: Mouton de Gruyter.
Talmy, Leonard
    1985   Force dynamics in language and cognition. *Cognitive Science* 12: 49-100.
    2000   *Toward a Cognitive Semantics*. Vol. 1. *Concept structuring systems*. Cambridge, MA: MIT Press.
    2008   Foreword. In *Methods in Cognitive Linguistics*, Monica Gonzalez-Marquez, Irene Mittleberg, Seanna Coulson and Michael Spivey (eds.), xxi – xxi. Amsterdam / Philadelphia: John Benjamins.
Tomasello, Michael
    2000   First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 11: 61–82.
Tummers, José, Dirk Speelman and Dirk Geeraerts
    2005   Inflectional variation in Belgian and Netherlandic Dutch: A usage-based account of the adjectival inflection. In *Perspectives on Variation. Sociolinguistic, historical, comparative*, Nicole Delbecque, Johan van der Auwera and Dirk Geeraerts (eds.), 93-110. Berlin / New York: Mouton de Gruyter.
Tummers, Jose, 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.

Wiechmann, Daniel

2008a   On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory* 4: 253-290.

2008b   Sense-contingent lexical preferences and early parsing decisions. *Cognitive Linguistics* 19: 439-455.

Wierzbicka, Anna

1985   *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.

1990   Prototypes 'save': on the uses and abuses of the notion of 'prototype' in linguistics and related fields. In *Meanings and Prototypes. Studies on linguistic categorization*, Savas Tsohatzidis (ed.), 347-367. London: Routledge.

Wulff, Stefanie

2003   A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8: 245-82.

2006   *Go*-V vs. *go-and*-V in English: A case of constructional synonymy? In *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*, Stefan Th. Gries and Anatol Stefanowitsch (eds.), 101-126. Berlin / New York: Mouton de Gruyter.

2009a   *Rethinking Idiomaticity. A usage-based approach*. London: Continuum.

2009b   Converging evidence from corpus and experimental data to capture idiomaticity. *Corpus Linguistics and Linguistics Theory* 5: 131-159.

Wulff, Stefanie, Anatol Stefanowitsch and Stefan Th. Gries

2007   Brutal Brits and persuasive Americans: Variety-specific meaning construction in the *into*-causative. In *Aspects of Meaning Construction*, Günter Radden, Klaus-Michael Köpcke, Thomas Berg and Peter Siemund (eds.), 265-281. Amsterdam / Philadelphia: John Benjamins.

Zeschel, Arne (ed.)

2008   *Usage-Based Approaches to Language Processing and Representation* (Special ed. *Cognitive Linguistics* 19). Berlin / New York: Mouton de Gruyter.

Zeschel, Arne

2009   What's (in) a construction? In *New directions in Cognitive Linguistics*, Vyvyan Evans and Stéphanie Pourcel (eds.), 185 - 200. Amsterdam / Philadelphia: John Benjamins.