UNIVERSITYᵒᶠ BIRMINGHAM

University of Birmingham
Research at Birmingham

Four challenges for usage-based linguistics.
Divjak, Dagmar

Document Version
Peer reviewed version

Citation for published version (Harvard):
Divjak, D 2015, Four challenges for usage-based linguistics. in J Daems, E Zenner, K Heylen, D Speelman & H Cuyckens (eds), *Change of Paradigms: New Paradoxes. : Recontextualizing Language and Linguistics..* Applications of Cognitive Linguistics , vol. 31, De Gruyter, Berlin, pp. 297-309.

Link to publication on Research at Birmingham portal

Four challenges for usage-based linguistics[1]

Dagmar Divjak

The University of Sheffield

Abstract

Dirk Geeraerts has played a key role in launching Cognitive Linguistics as a full-fledged theory of linguistics and in expanding its sphere of influence in Western Europe. Dirk is furthermore one of the first and strongest advocates for the incorporation of empirical methods - and quantitative, corpus-based methods in particular - into cognitive linguistic research. The "Quantitative Turn" (Janda 2013) is in large part due to his relentless insistence on methodological rigour. In this chapter, I want to take a closer look at what is currently methodological "good practice" in the field and draw attention to some of the assumptions that underlie our methodology and thereby shape our findings yet have gone unquestioned. Four challenges are highlighted - data annotation, statistical analysis, model validation and experimental design - and their theoretical foundations and implications discussed.

Keywords: cognitive commitment; cognitive reality; corpus-linguistics; psycho-linguistics; statistical modelling; usage-based linguistics;

After half a century of self-imposed exile from the cognitive scene, cognitive linguists have put language back on stage — language is no longer considered a highly specialized and largely autonomous cognitive module that needs "special treatment". Instead, linguistic abilities are seen as rooted in general cognitive abilities and meaning is understood as conceptualization. In fact, cognitive linguists are bound by two major commitments: the generalization commitment and the cognitive commitment (Lakoff 1990: 40). All cognitive linguists are committed (or are assumed to be committed) to providing a characterization of the general principles governing all aspects of human language in a way that is informed by and accords with what is known about the brain and mind from other disciplines. Work in the Cognitive Linguistic tradition therefore likes to stress that the analyses proposed are "in line with what is known about the mind" and abounds with claims that the proposed analysis would be cognitively realistic, if not cognitively real. But is this really so?

Unlike many other modern theories of linguistics, Cognitive Linguistics also aims to be a usage-based model of language structure (Langacker 1987: 46). All language units arise from and are shaped by usage events by means of the aforementioned general cognitive abilities such as perception, attention, memory, categorization and abstraction. Usage events are observable, and therefore they can be collected, measured, and analyzed (Glynn 2010: 5–6). A decade ago, Tummers et al. (2005: 225-226) concluded that "[c]orpus linguistics would be an obvious methodology for a usage-based linguistics: you cannot have a usage-based linguistics unless you study actual usage - as it appears in an online and elicited form in experimental settings or as it appears in its most natural form in corpora in the shape of spontaneous, non-elicited language data". While in 2005 "the use of corpus materials [was] not yet the dominant approach, and to the extent that the research is actually corpus-based, a tendency toward the use of advanced corpus techniques [was] only beginning to emerge" (Tummers et al. 2005: 248), the situation is rather different now. In order to describe a phenomenon and uncover the mechanisms that govern it, linguists tend to turn to the linguistic analysis and statistical modeling of data from large corpora or elicited through experiments. Anno 2015, there are plenty of published articles that rely on data extracted from corpora and annotated for a multitude of morphological, syntactic, semantic and pragmatic parameters, to model a phenomenon and/or predict the choice for one morpheme, lexeme or construction over another. According to Janda (2013: 4) we can "divide the history of Cognitive Linguistics [the journal - D.D.] into two eras, 1990–2007 – when most articles were not quantitative, and 2008–2012 – when most articles were quantitative" [a "quantitative article" being defined as an article in which a researcher reports numbers for some kind of authentic language data]. She continues "[w]e can [...] securely identify 2008–2012 as a distinct period in the history of Cognitive Linguistics. During this period quantitative analysis emerges as common practice, dominating the pages of our journal" (Janda 2013: 6).

Dirk Geeraerts was instrumental in launching Cognitive Linguistics as a full-fledged theory of linguistics and has shown particular concern for its methodological machinery. In the next two sections I will take a closer look at the usage-based approach of which Dirk has been one of the strongest advocates, and I will discuss some of the challenges that this changed paradigm is currently facing. I will discuss the following challenges in turn:

Challenge 1. We work in a corpus-based fashion, at the heart of which lies the manual annotation of data. Do we reflect sufficiently on how our very first decisions affect our findings?

Challenge 2. We analyze our data statistically, using approaches from the frequentist tradition. Do we give enough consideration to the assumptions on which these techniques are based and to the implications that has for our findings?

Challenge 3. We capture human behavior in models, knowing that "all models are wrong" (Box 1976: 792). Are we sufficiently concerned about testing our models against human behavior?

Challenge 4. We run experiments on language, complying with methodological requirements developed for other aspects of human behavior. Should we not pause to consider whether the nature of language meshes with the standard designs?

In other words, I want to draw attention to assumptions that underlie our choice of methods and hence shape our findings, yet have hitherto gone unquestioned.

Challenge 1: data annotation categories and principles vary widely

An important contribution to the statistical analysis of linguistic data -- of any data, really -- is made by the variables used to capture the phenomenon. At the heart of a corpus-based study of linguistic phenomena lies the (often manual) annotation of examples. These data annotations are typically "linguistic" in nature, that is, they are based on categories that were designed to aid the description of a language's form and meaning. Some of these categories have been around for millennia; the classification of words into categories, for example, predates Christianity. As early as the 5th century BC, Sanskrit grammarians grouped words into classes – that would later become known as parts of speech – distinguishing between inflected nouns and verbs and uninflected pre-verbs and particles. Other linguistic categories that are well established in theoretical linguistics, regardless of framework, are, for example, phonemes, morphemes, tense, mood, aspect etc. Cognitive Linguistics has created its own categories, such as image schemas, trajectors and landmarks, conceptual metaphors, constructions and frames. With few exceptions the universality of the adopted traditional linguistic categories has gone unquestioned (e.g. Evans & Levinson 2009) and the cognitive reality of the newly introduced cognitive linguistic categories has not been systematically addressed (cf. Gibbs & Colston 1995).

Linguistic reality and psychological reality seem to have become one, resulting in a situation whereby linguists elevate linguistic descriptions to psychological explanations and psycholinguists expect to find evidence of the cognitive reality of classifications that were designed to aid the description of language data, not to reflect the workings of the mind (compare also Eddington 2002: 212-213). Yet "[c]ognitively real generalizations may not at all accord with generalizations arrived at by classical techniques of linguistic analysis" (Lakoff 1990: 41). In fact, there is no agreed-upon definition of what is meant by "cognitively real(istic)" and what level of cognitive commitment is expected. Categories that are "consistent with our overall knowledge about cognition and the brain" (Lakoff 1990: 45) could well range from categories that can be presented as radial categories with prototype structure to those for which there is neurological evidence, i.e. a unique neurological signature that proves that a category is treated as a processing unit in its own right by actual language users.

A second question that would benefit from more consideration relates to the nature and extent of our data annotation: a typical analysis involves coding a large number of extractions for a number of properties, yet studies diverge in their implementation of this principle. The vast majority of studies annotate their data for a limited number of properties that operationalize a specific hypothesis. Some more recent studies, however, explicitly advocate the annotation of as many potentially relevant properties as possible in as linguistically naive a way as possible (Arppe 2008, Divjak 2010). While the former approach seems suited if we aim to pitch competing linguistic hypotheses against each other, the latter approach is more appropriate if we are interested in letting the relevant patterns fall out from the data (but see Challenge 2). In fact, Cognitive Linguistics has been "accused" of using "categories gained from introspection rather than from the data itself" (Teubert 2005: 2). Syntactic, semantic and discourse-related higher-level abstract features are believed to help reveal more general patterns (Theijssen et al. 2013: 228) but recent research has shown that including these features - that are often difficult to define and to annotate with high agreement levels between human annotators - does not necessarily yield a better model than working with lexical features, such as the actual words used (Theijssen et al. 2013: 246, 257). An approach that stays close to the raw data and captures "every possible clue" comes with the added benefit that "[k]eeping as much detail[ed] information as long as possible — even throughout advanced stages of analysis — is crucial because we never know if what we believe to be the relevant features really are the only essential ones" (Wälchli & Cysouw 2012:703).

Challenge 2: "probabilistic" is a polysemous word

Usage constitutes the dataset in which general patterns can be detected, and this is more and more frequently done by making use of statistical techniques. Reliance on data and statistics certainly gives us more confidence in our conclusions, but does it guarantee that our models are any cognitively more real(istic) than they were before? We do not seem to worry very much about detecting patterns in a cognitively realistic fashion. Much of modern statistics was developed on the basis of the frequentist (rather than Bayesian) interpretation of probability and we readily adopt frequentist techniques to model our data. For those models to be cognitively real(istic), we would need to assume that probabilistic reasoning underlies language knowledge and use. But, "probabilistic" is a polysemous word and in linguistic circles, the non-technical meaning of "supported by evidence strong enough to establish presumption but not proof" appears to prevail. Probabilistic grammars are seen as opposed to rule-based grammars and this reflects the insight that the phenomenon studied is not fully predictable. As Kilgariff (2005) and many others have observed: language is never ever random; however, it is also rarely, if ever, fully predictable.

This "linguistic" interpretation of the statistical term "probabilistic" is rather different from the frequentist statistical interpretation as "the ratio of the number of outcomes in an exhaustive set of equally likely outcomes that produce a given event to the total number of possible outcomes" or "the chance that a given event will occur". According to frequentists, the probability of an event is defined as the relative frequency of the event in some reference class (Lassiter 2012). The reference class is a core component in the probability calculation and one that is highly problematic: it makes the probability of an event dependent on the choice of a reference class, and because an event belongs to many reference classes, it is not always obvious which reference class to choose (cf. the cell 4 problem reported for

collostructional analysis in Schmid & Küchenhoffer 2013). Moreover, the interpretation of probability as relative frequency cannot make intuitive sense of the fact that probabilities can attach to non-repeatable events (Lassiter 2012): according to the frequentist definition, the probability of an event that can only happen once is either 1 (if it happens) or 0 (if it does not happen). A variant of frequentism (von Mises 1957) therefore claims that the probability of an event should be identified with the relative frequency in a hypothetical sequence generated by flipping the coin an infinite number of times.

A more palatable approach to uncertainty is found in Bayesianism, which remains rare in linguistics, however. For Bayesians, probability is weight of evidence: it is a measure of a rational agent's degree of belief in a proposition (Lassiter 2012). Bayesian methods apply in a wider range of situations than frequentist methods, and are more flexible. Crucially, Bayesian methods can be applied to estimating probabilities for repeatable and non-repeatable events and it is possible to incorporate prior information into a model (Lassiter 2012). This seems crucial for modeling cognitive phenomena such as language, since human beings usually approach inference problems with some prior knowledge. However, research from decision making has shown that people have extreme difficulty if information is given and answers are asked for in single-event probabilities; but they appear to behave like good "intuitive statisticians" when information is given and answers are asked for in frequencies (Brase et al. 1998: 19).

Challenge 3: models are rarely tested on speakers

The number of publications relying on empirical data collections and statistical data modelling has increased spectacularly. The most advanced analyses rely on regression analyses to model which of the candidate properties are predictive of the form which is the focus of the study. These techniques are attractive because they allow to (1) estimate the relative weights of the linguistic explanatory variables in natural terms as odds, and to (2) model the impact of the co-occurrence of these variables in various combinations as expected probability distributions for the alternative lexemes/constructions. Yet neither frequentist nor Bayesian models are based on learning mechanisms.

If we want our linguistics to be cognitively realistic, should we not consider using modeling techniques that are directly based on principles of human learning? Several models of learning have been implemented for and tested on language data, and the predictions have been compared to the behavior of subjects in experimental settings. The best-known ones in Cognitive Linguistic circles are Connectionist Modelling (Rumelhart & McClelland 1986), Analogical Modelling (Skousen 1989), Memory-based Learning (Daelemans & van den Bosch 2005) and more recently Naive Discriminative Learning (Baayen 2010). Eddington (2000) compared a connectionist model with an analogical model and a memory-based model in their performance on the English past tense. His findings showed that, different from the connectionist model that only handled the irregular items well, the analogical and memory-based models successfully predicted subject's choice of past tense for nonce verbs for both regular and irregular items, and they did so by comparing the nonce words to words in the database in terms of their phonological similarity. Theijssen et al (2013) compared the performance of logistic regression (using higher-level features), Bayesian networks (using higher-level features) and Memory-based learning (using lexical items) in predicting the English dative alternation. They found the overall performance of the three models to be virtually identical, although the classification of the individual

cases by the memory-based model differed most from the other two approaches. Baayen et al. (2013) have shown that statistical classifiers based on cognitively realistic approximations of how humans learn such as NDL perform as well as regression models for binary choices. Preliminary results support this finding for more complex corpus models that predict a 4-way polytomous choice (Arppe & Baayen 2011).

Another important challenge faced by linguists is the question of how to evaluate such models. The most rigorous studies fit a statistical model to one part of the data (the training set) and test it on a new set of corpus examples (the testing set) to see how well the findings generalize to new data. But is a corpus-based model with high predictive power satisfactory even if the model's performance is not tested against speakers' performance? If interest is in modelling *human* knowledge, should we not compare our models' performance to that of native speakers of the language? Surprisingly few papers currently attempt this (for an overview, see Klavan & Divjak, under review) and linguists who run an experimental study after a corpus-based study often refer to this process as "validation". This, unfortunately, creates the impression that behavioral experimental data is inherently more valuable than textual data, be it transcribed spoken language or originally written language. But for language, textual data is the result of one of the most natural types of linguistic behavior: "[a] corpus is a collection of non-elicited usage events. It constitutes a sample of spontaneous language use that is (generally) realized by native speakers" (Tummers et al. 2005: 231). Observing the output qualifies as an "observational study" and possibly as a "natural experiment"; these types of experiments are quite popular in disciplines where experimental manipulation of groups and treatments would be unethical, e.g. epidemiology. Through observation we get a real picture of the phenomenon as it manifests in natural settings, although we should not forget that corpus data is not actually representative of any single speaker; instead, it represents a non-existing average speaker (cf. Blumenthal-Dramé 2012: 30, 34).

Challenge 4: language in the lab versus language in use

Experiments, and laboratory-based experiments in particular, afford the researcher a high level of control over variables; by manipulating the variables, it becomes possible to establish cause and effect relationships. Due to the need to maintain control over the variables experimental studies are often run in artificial settings. It is maintained that the physical situations in the real world and in the lab may differ, provided that the same processes are occurring. And this is where the shoe pinches: experimental linguistic studies standardly present words in isolation or use artificially constructed stimuli that bear little resemblance to naturally produced data. There are at least two reasons to suspect that the customary approach to stimulus selection makes it unlikely that the same processes that regulate language use in real life occur in the lab.

First, when selecting stimuli for an experiment, psycholinguists do not routinely conduct a multivariate corpus-based analysis, and often limit their interest in corpora to the possibility to use them as source of information on the frequency of occurrence of words or chunks. This is because frequency, just like familiarity and length, is known to exert a strong influence on a number of behaviors, including

processing speed: to avoid the "confound of frequency" when comparing reaction times to different categories of words or other language structures experimental items are routinely matched for frequency. Yet extracting only frequency of occurrence information from a corpus severely impoverishes the richness of the linguistic experience from which learners extract patterns; there is more to a word than its frequency of occurrence. In fact, in a natural setting so many factors influence a given phenomenon that any selection, not based on an exhaustive, i.e. multivariate, study of the phenomenon is a stab in the dark. Speakers have extremely specific expectations about words that are learned from encountering those words in their natural contexts. It has been shown that probabilities are essential for a cognitive model of sentence processing (Jurafsky 1996) and Divjak & Arppe (2014) have established that this is not only the case when probable combinations are compared to (artificially created) improbable combinations but even when all combinations occur naturally, i.e. are *more or less* likely. Context is another confound, yet one that is routinely ignored. Psycholinguists might want to worry less about length-based, familiarity-based, or frequency-based lexical effects and more about properties of language in use that might affect experimental results. Ignoring the dependence of a word and the specific form it occurs in on its context may well skew our understanding of the cognitive mechanisms underlying word processing.

Second, experimental settings may "force participants to tackle problems that are not faced in normal discourse" (Deignan 2005: 117). While it is justified to adhere to standard experimental methodology in order to identify basic mechanisms such as frequency effects that are of core relevance to the theory, it seems questionable to also apply these methods when validating specific predictions. The results of the latter type of experiments may well tell us something interesting about the processing of X under condition Y, but if condition Y is not typically encountered in reality, they do not tell us much about the processing of language in use. Now that corpus-based techniques are available to calculate the probability of a word (form) given all other words in the sentence, and advances in analysis techniques make it possible to control for a plethora of factors statistically, running experiments with words in their natural contexts is achievable. This will, in fact, bring closer the ideal of "controlling everything but the variables that are being manipulated" while also ensuring the external validity of the findings. The ecological validity of laboratory results has been questioned in more general terms. Mitchell (2012) aggregated results of several meta-analyses and concluded that, although many psychological results found in the laboratory can be replicated in the field, their effects often differ greatly in size and sometimes even in direction (Mitchell 2012: 114). It remains to be seen to what extent current experimental linguistic findings are side-effects of the experimental settings used.

Where do we go from here?

Paraphrasing Divjak (2012) I conclude that studying language in use is a discipline at the intersection of linguistics and psychology. Yet many psycholinguistic studies have been carried out by research teams that do not include (corpus) linguists who love getting their hands dirty in the data. Counting readily identifiable forms taken out of their natural context significantly diminishes the richness of the input from which human beings extract and learn distributional patterns. At the same time, many cognitive corpus linguistic studies continue to take their painstakingly annotated textual datasets to be a pretty

reliable map of speakers' minds, forgetting that what is learned or acquired by probabilistic means is not strictly proportional to the stimulus (and that frequency of occurrence is not the "be all and end all" in language, see Baayen 2010; Ellis 2012). Probabilistic learning theory holds that language learning is based on complex, higher-order properties of probabilistic patterns in sensory experience, not a mere tabulation of frequency of patterns (Elman 2003). Driven to its extreme, this split approach reduces our billion-neuron brains that enable us to adapt quickly to an immense array of stimuli to nothing more than sophisticated abacuses used to keep tallies of all the words found in the messy bag that language is.

David Poeppel tweeted recently (05.01.2015) "I'm pretty tired of big data and definitely ready for big theory. Let's stop collecting so much damn data and use 2015 to think about stuff". It wouldn't be a bad idea to take a break from collecting and modeling data and indeed spend some time "thinking about stuff", about the methodological questions raised in this chapter, and about their theoretical foundations and implications. Twenty five years ago Lakoff (1990: 43, 36) wrote that "I am sure that others who consider themselves cognitive linguists do not have the same primary commitments that I do, and that disagreements over how to properly analyze a given phenomenon are sure to follow from differences in primary commitments. [...] Without agreement on initial premises, arguments about conclusions will be pointless". That is as true today as it was then. Among the questions we need to answer are the following: are we, or are we not, concerned with cognitively real generalizations? What do we mean by "cognitively real generalizations"? If we take cognitively real generalizations to encompass only that for which evidence can be found in the minds of speakers, can we, or can we not, arrive at such generalizations given that our data elicitation paradigms do not require that language is studied in use; our data annotation schemas hinge on linguistic insights that the average speaker may well lack; and our modeling techniques are not implementations of the way in which human beings learn? Lakoff (1990: 41) pointed out, "If we are fortunate, these [i.e. generalization and cognitive] commitments will mesh: the general principles we seek will be cognitively real. If not, the cognitive commitment takes priority: we are concerned with cognitively real generalizations. This is anything but a trivial matter". That, too, is as true today as it was then.

References

Allen, Kachina, Francisco Pereira, Matthew Botvinick and Adele E. Goldberg. 2012. Distinguishing Grammatical Constructions with fMRI Pattern Analysis. *Brain and Language* 123: 174-182.

Arppe, Antti. 2008. *Univariate, bivariate and multivariate methods in corpus-based lexicography – a study of synonymy*. Publications of the Department of General Linguistics, University of Helsinki, No. 44. URN: http://urn.fi/URN:ISBN:978-952-10-5175-3.

Antti Arppe and R. Harald Baayen. 2011. *Statistical classification and principles of human learning.* 4th Conference on Quantitative Investigations in Theoretical Linguistics (QITL4), Humboldt-Universität zu Berlin, Germany, 28-31 March 2011.

Baayen, R. Harald. 2010. Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon* 5, 436-461.

Baayen, R. Harald., Laura A. Janda, Tore Nesset, Stephen Dickey, Anna Endresen and Anastasia Makarova. 2013. Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37: 253-291.

Blumenthal-Dramé, Alice. 2012. *Entrenchment in Usage-Based Theories. What Corpus Data Do and Do Not Reveal About The Mind*. Series:Topics in English Linguistics 83.
Box, George E. P. 1976. Science and Statistics. *Journal of the American Statistical Association*, 71 (356): 791-799.

Brase, Gary L., Leda Cosmides and John Tooby. 1998. Individuation, counting, and statistical inference: The role of frequency and whole-object representations in judgment under uncertainty. *Journal of Experimental Psychology General* 127(1): 3-21.

Christensen, K. R. and Wallentin, M. 2011. The locative alternation: distinguishing linguistic processing cost from error signals in Broca's region. *NeuroImage*, 56 (3), 1622-31.

Daelemans, Walter and Antal van den Bosch. 2005. *Memory-Based Language Processing.* Cambridge, UK: Cambridge University Press.

Deignan Alice H. 2005. *Metaphor and Corpus Linguistics.* Amsterdam: John Benjamins.

Divjak, Dagmar. 2010. *Structuring the Lexicon: a Clustered Model for Near-Synonymy.* Berlin: De Gruyter. [Cognitive Linguistics Research 43].

Divjak, Dagmar. 2012. Introduction to Divjak, D. & Gries, St. Th. 2012. *Frequency effects in language representation (Vol. 2).* Mouton de Gruyter (Berlin – New York), 1-10. [Trends in Linguistics].

Divjak, Dagmar and Antti Arppe 2013. Extracting prototypes from exemplars. What can corpus data tell us about concept representation? *Cognitive Linguistics*, 24 (2): 221-274.

Divjak, Dagmar and Antti Arppe 2014. *Unusual suspects: TAM markers predict lexical choice best.* Oral presentation at UK-CLC 5, 29-31 July, Lancaster (UK).

Divjak, Dagmar and Stefan Th. Gries (eds). 2012. *Frequency Effects in Language Representation*. Berlin/New York: Mouton de Gruyter.

Eddington, David. 2000. Analogy and the dual-route model of morphology. *Lingua* 110: 281-298.

Eddington, David. 2002. Why Quantitative? *Linguistics* 40: 209-216.

Ellis, Nick. 2012. What can we count in language, and what counts in language acquisition, cognition, and use? In Gries & Divjak *Frequency effects in language acquisition and use*. De Gruyter Mouton: Berlin. [Trends in Linguistics 244.1]

Elman, J. L. 2003. Generalization from sparse input. In *Proceedings of the 38th Annual Meeting of the Chicago Linguistic Society.*

Evans, Nicholas and Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5): 429-492.

Gibbs Jr., Raymond W. and Herbert L. Colston. 1995. The cognitive psychological reality of image schemas and their transformations. *Cognitive Linguistics* 6(4): 347-378.

Glynn, Dylan and Kerstin Fischer. (eds) 2010. *Corpus Cognitive Semantics. Usage-based methods for the study of meaning.* Berlin: Mouton de Gruyter.

Gries, Stefan Th. and Dagmar Divjak (eds.). 2012. *Frequency Effects in Language Learning and Processing*. Berlin/New York: Mouton de Gruyter.

Janda, Laura (ed). 2013. *Cognitive Linguistics: The Quantitative Turn. The Essential Reader*. Berlin: De Gruyter Mouton.

Jurafsky, Dan. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science,* 20, 137–194.

Kilgariff, Adam. 2005. Language is never ever ever random. *Corpus Linguistics and Linguistic Theory* 1-2: 263-276.

Klavan, Jane and Dagmar Divjak. (Under review). Review article. The Cognitive Plausibility of Statistical Classification Models: Comparing Textual and Behavioral Evidence. *Folia Linguistica.*

Lakoff, George. 1990. The Invariance Hypothesis: is abstract reason based on image-schemas? *Cognitive Linguistics* 1(1): 39 - 74.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Theoretical Prerequisites.* Stanford, CA: Stanford University Press.

Lassiter, D. 2012. Probabilistic reasoning and statistical inference: An introduction (for linguists and philosophers). [http://web.stanford.edu/~danlass/NASSLLI-coursenotes-combined.pdf; last accessed 21/2/2015].

von Mises, Richard. 1957. *Probability, Statistics and Truth.* 2nd English edition. George Allen & Unwin.

Poeppel, David. 2015, January 5. I'm pretty tired of big data and definitely ready for big theory. Let's stop collecting so much damn data and use 2015 to think about stuff. [Twitter post]. Retrieved from [https://twitter.com/davidpoeppel/status/552202650834731008].

Rumelhart,D. E. and J. L. McClelland. 1986. On learning the past tense of English verbs. In: J.L. McClelland, D.E. Rumelhart, PDP Research Group (eds.), *Parallel distributed processing,* vol. 2, 216-271. Cambridge, MA:MIT Press.

Schmid, Hans-Jörg and Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3), 531-577.

Skousen, Royal. 1989. *Analogical modeling of language*. Dordrecht: Kluwer Academic.

Teubert, Wolfgang. 2005. My version of corpus linguistics. *International Journal of Corpus Linguistics*, 10 (1), 1–13.

Theijssen, Daphne, Louis ten Bosch, Lou Boves, Bert Cranen and Hans van Halteren. 2013. Choosing alternatives: Using Bayesian Networks and memory-based learning to study the dative alternation. *Corpus Linguistics and Linguistic Theory* 9 (2), 227–262.

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.

Wälchli, Bernhard and Michael Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50 (3): 671–710.