Usage-Based Cognitive Semantics: A Quantitative Approach

Dylan Glynn & Kerstin Fischer
University of Leuven & University of Hamburg

Over the last two decades, Cognitive Linguistics has seen a slow but sure move towards data-driven methodology. This theme session is designed to promote the development of such methodology in Cognitive Semantic research. Building on the usage-based theme sessions of previous ICLCs, our focus is the quantitative treatment of lexical and constructional semantics.

Although we do not question the importance of elicitation, experimentation, and intuition, this theme session focuses on the use of found data. Corpus data respect the complexity of language and, if treated in sufficiently large quantities, enable generalisations about language structure that other methods cannot. This is true not only for the study of form-meaning pairs but also for understanding the interaction between different types of formal structure, such as lexis and syntax. An important feature of usage-based quantitative methodology is that it facilitates attempts to reveal the interaction between the different parameters of language simultaneously. Such a method allows us to capture the multifactorial effects of register, dialect, and context on language usage, parameters integral to semantic structure. Finally, quantitative methods should be brought to the fore since they offer the best means for result verification.

Early cognitive semantic research, such as Dirven & al. (1982) and Lehrer (1982), pathed the way for the development of quantitative techniques in lexical research. Geeraerts & al. (1994), Schmid (2000), and Fischer (2000) are exemplary of this movement. Their research focuses on both polysemy and parasynonymy in lexical structure. More recently, corpus-driven techniques have been applied to constructional semantics, for example Gries (2003) and Stefanowitsch (2003). This research has focused on lexical-constructional licensing and syntactic alternation or parasynonymy. The ensemble of such approaches has been brought together in recent anthologies, Gries & Stefanowitsch (2006), Stefanowitsch & Gries (2006), and Newman & Rice (in press).

We assume that the study of lexical semantic structures necessarily includes their collocational behaviour, just as the study of constructional semantics necessarily includes the study of lexical licensing. Moreover, both of these endeavours should be concerned with extra-linguistic factors as they surface in register and dialect and the role of contextual information. This theme session hopes to bring these analytical concerns together through the development of corpus-driven methodology, in particular issues of annotation, corpus representativity, and statistical analysis.

Dirven, R. & al. 1982. *The scene of linguistic action and its perspectivization by SPEAK, TALK, SAY and TELL*. Benjamins: Amsterdam.

Fischer, K. 2000. *From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Particles*. Mouton: Berlin.

Geeraerts, D. & al. 1994. *The Structure of Lexical Variation. Naming, meaning, and context*. Mouton: Berlin.

Gries, St. Th. 2003. *Multifactorial Analysis in Corpus Linguistics*. Continuum: London.

Gries, St. Th. & Stefanowitsch, A. (eds). 2006. *Corpora in Cognitive Linguistics*. Mouton: Berlin.

Lehrer, A. 1982. *Wine and Conversation*. IUP: Bloomington.

Newman, J. & Sally R. (eds). In press. *Empirical and Experimental Methods in Cognitive/Functional Research*. CSLI: Stanford.

Schmid, H.J. 2000. *English Abstract Nouns as Conceptual Shells. From corpus to cognition*. Mouton: Berlin.

Stefanowitsch, A. 2003. Constructional semantics as a limit to grammatical alternation. G. Rohdenburg & B. Mondorf (eds), *Determinants of Grammatical Variation in English*. Mouton: Berlin.

Stefanowitsch, A. & Gries, St. Th. (eds). 2006. *Corpus-based Approaches to Metaphor and Metonymy*. Mouton: Berlin.

Theme Session Structure

The papers fall into 4 basic sub-themes:

i. Collostructional Methods
ii. Methods in Construction Grammar
iii. Semantics and Corpus Driven research
iv. Representativity and Statistical Techniques

Each of the four sub-themes is briefly introduced and the important questions for that sub-theme are outlined. The participants speak without pauses for questions, save for points of clarification. Each thematic set of speakers is followed by about 20 minutes of discussion and questions. We propose to allocate five slots for the panel, three 2-hour and two 90-minutes slots, which will ideally be distributed such that the two theoretical sessions on issues of representativity frame the more empirically-oriented sessions.

Representativity and Statistical Techniques I
Balancing Acts: Empirical Pursuits in Cognitive Linguistics
(John Newman)
Corpus-based preposition studies: the case of Palikur
(Pierre Cadiot & François Nemo)
Does frequency in text really instantiate entrenchment in the cognitive system?
(Hans-Jörg Schmid)

Collostructional Methods
On the futurate present in English and German
(Martin Hilpert)
Beyond the dative alternation: the semantics of the Dutch prepositional-dative construction
(Timothy Colleman)
Analyzing constructional pragmatics: a usage-based study of German word order constructions.
(Kris Heylen)
A collostructional analysis of the causative alternation in English
(Maarten Lemmens)

Semantics and Corpus Driven research
Comparable Corpus Elicitation for Cognitive Semantic Analysis
(Kerstin Fischer)
The Semantics of Language Use: Why Corpus Matters
(François Nemo)
Quantifying meaning: where do the English and Russians start/begin?
(Stefan Th. Gries & Dagmar Divjak)

Methods in Construction Grammar
Marrying cognitive-linguistic theory and corpus-based methods: On the compositionality of English V-NP idioms
(Stefanie Wulff)
A usage-based approach to constructional ambiguity: the case of Dutch participles
(Gert De Sutter)
Quantifying the interaction between modality and aspect in Slavic languages
(Dagmar Divjak & Oesten Dahl)
Information structure and constructional semantics
(Anatol Stefanowitsch)

Representativity and Statistical Techniques II
Exemplars and analogy: semantic extension in constructional networks
(Arne Zeschel)
The cognitive reality of frequent verb-noun combinations: An empirical study
(Gaëtanelle Gilquin)
Testing the hypothesis. Confirmatory statistical techniques for multifactorial data in Cognitive Semantics
(Dylan Glynn, Dirk Geeraerts, & Dirk Speelman)

Participants, Affiliations, and Paper Titles
(contact for first author)

Cadiot, Pierre & Nemo, François
CORAL, University of Orleans <francois.nemo@univ-orleans.fr>
Corpus-based preposition studies: the case of Palikur (Arawakan)

Colleman, Timothy
Ghent University <timothy.colleman@ugent.be>
Beyond the dative alternation: the semantics of the Dutch prepositional-dative construction

De Sutter, Gert
Ghent University <gert.desutter@arts.kuleuven.be>
A usage-based approach to constructional ambiguity: the case of Dutch participles

Divjak, Dagmar & Dahl, Östen
University of Sheffield, Belgian Science Foundation & Stockholm University <d.divjak@sheffield.ac.uk>
Quantifying the interaction between modality and aspect in Slavic languages

Fischer, Kerstin
University of Hamburg <fischer@nats.informatik.uni-hamburg.de>
Comparable Corpus Elicitation for Cognitive Semantic Analysis

Gilquin, Gaëtanelle
FNRS, Université catholique de Louvain <gilquin@lige.ucl.ac.be>
The cognitive reality of frequent verb-noun combinations: An empirical study

Glynn, Dylan
QLVL, University of Leuven <dsglynn@paris7.jussieu.fr>
Operationalising polysemy. A Quantitative Multifactorial study in Cognitive Semantics

Glynn, Dylan, Geeraerts, Dirk, & Speelman, Dirk
QLVL, University of Leuven <dsglynn@paris7.jussieu.fr>
Testing the hypothesis. Confirmatory statistical techniques for multifactorial data in Cognitive Semantics

Gries, Stefan Th. & Divjak, Dagmar
University of California, Santa Barbara & University of Sheffield <stgries@linguistics.ucsb.edu>
Quantifying meaning: where do the English and Russians start/begin?

Heylen, Kris
QLVL, University of Leuven <kris.heylen@arts.kuleuven.be>
Analyzing constructional pragmatics: a usage-based study of German word order constructions.

Hilpert, Martin
Rice University <hilpert@rice.edu>
On the futurate present in English and German

Lemmens, Maarten
Université Lille3 & CNRS <maarten.lemmens@univ-lille3.fr>
A collostructional analysis of the causative alternation in English

Nemo, François
University of Orleans <francois.nemo@univ-orleans.fr>
The Semantics of Language Use: Why Corpus Matters

Newman, John
University of Alberta <john.newman@ualberta.ca>
Balancing Acts: Empirical Pursuits in Cognitive Linguistics

Schmid, Hans-Jörg
University of Munich <hans-joerg.schmid@anglistik.uni-muenchen.de>
Does frequency in text really instantiate entrenchment in the cognitive system?

Stefanowitsch, Anatol
University of Bremen <stefanowitsch@uni-bremen.de>
Information structure and constructional semantics

Wulff, Stefanie
University of Bremen <stefaniewulff@gmail.com>
Marrying cognitive-linguistic theory and corpus-based methods: On the compositionality of English V-NP idioms

Zeschel, Arne
University of Bremen <zeschel@uni-bremen.de>
Exemplars and analogy: semantic extension in constructional networks

Email list of all particpants:
zeschel@uni-bremen.de, stefaniewulff@gmail.com, stefanowitsch@uni-bremen.de, hans-joerg.schmid@anglistik.uni-muenchen.de, john.newman@ualberta.ca, francois.nemo@univ-orleans.fr, maarten.lemmens@univ-lille3.fr, hilpert@rice.edu, stgries@linguistics.ucsb.edu, dsglynn@paris7.jussieu.fr, fischer@nats.informatik.uni-hamburg.de, gilquin@lige.ucl.ac.be, d.divjak@sheffield.ac.uk, francois.nemo@univ-orleans.fr, gert.desutter@arts.kuleuven.be, timothy.colleman@ugent.be, oesten@ling.su.se, cadiot@ens.fr, dirk.geeraerts@arts.kuleuven.be, dirk.speelman@arts.kuleuven.be, kris.heylen@arts.kuleuven.be

Corpus-based preposition studies: the case of Palikur (Arawakan)

Pierre Cadiot & François Nemo
University of Orleans (CORAL)

In this paper, we compare the effects of using two different methodologies applied to the same data set, namely spatial prepositions. The study of prepositions in itself already constitutes a challenge for cognitive semantics since it involves the study of both the relationship between language and space (or language and time, etc.), and of the relationship between spatial and non-spatial uses (temporal versus non temporal uses, etc.). Furthermore, prepositions constitute a domain in which typological differences with respect to the lexicalization and grammaticalisation of the expression of space (or time, etc.) or polyfunctionality clearly forbid generalizations based on a single language.

It is in connection with these issues that we intend to contrast the results of two studies and surveys of prepositions in the Palikur (Arawakan, Maipuran) language of French Guyana and Brazil (Amapa), made by the same team but based on different methodologies. The first study uses classical techniques of investigation while the second is based on data extracted from a corpus.

Palikur is a language with a complex system of numeral classifiers that grammaticalises eight different geometrical shapes, in which prepositions often have a much more specific (and often dynamic) meaning than their French or English translations. Moreover, many of these items are polycategorial, showing so-called homonyms in different grammatical categories.

We shall specifically discuss the insights provided by the two types of data concerning:

- the grammatical/categorical status of the morphemes presumably labelled prepositions;
- the adequacy/inadequacy of a purely spatial characterization of their core meaning;
- the problematisation of isolating spatial uses among other uses;
- the possibility of adopting a distributional methodology that uses polyfunctionality and polycategoriality as a tool in order to isolate meaning;
- the possibility of integrating depictive, topological, praxeological, modal, and axiological dimensions of meaning.

Our study demonstrates the importance of methodology in linguistic analysis.

Cadiot, P. 1997. *Les prépositions abstraites en français*. Paris: Colin.

- 2001. *Pour une théorie des formes sémantiques : motifs, profils, thèmes.* Paris: PUF.
- 2002. Schematics and motifs in the semantics of prepositions. *Prepositions in their Syntactic, Semantic and Pragmatic Context*. S. Feigenbaum & D. Kurzon (eds), 41–57. Amsterdam: Benjamins.

Cadiot, P., Lebas, F. & Visetti, Y.-M. 2006. The semantics of motion verbs: Action, space, and qualia. *Space in languages: linguistic systems and cognitive categories*. M. Hickmann & S. Robert (eds), 175-206. Amsterdam: Benjamins

Launey Michel (2003). *Awna Parikwaki. Introduction à la langue palikur de Guyane et de l'Amapa*. Paris: IRD.

Nemo, F. 2004. Constructions et morphèmes : réflexions sur la stabilité en sémantique. *Revue de Sémantique et Pragmatique* 15: 19-35

Tyler, A. & Evans, V. 2003. *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition*. Cambridge: CUP.

Visetti, Y.-M., & Cadiot, P. 2002. Instability and the theory of semantic forms: Starting from the case of prepositions. *Prepositions in their Syntactic, Semantic and Pragmatic Context.* S. Feigenbaum & D. Kurzon (eds), 9–39. Amsterdam: Benjamins.

Beyond the dative alternation:
the semantics of the Dutch prepositional-dative construction

Timothy Colleman
Ghent University

One of the basic theoretical/methodological guidelines of Construction Grammar is that each construction should first and foremost be analysed in its own right (cf. Goldberg 2002, inter alia), a principle which is often ignored in studies of argument structure alternations (or other cases of "competing" constructions). The majority of existing accounts of the indirect object constructions of Dutch, for instance, aims at pinpointing a number of schematic semantic contrasts between the ditransitive construction in (1) and its prepositional "paraphrase" in (2), without providing elaborate accounts of the semantic ranges of both constructions.

(1) Jan heeft Piet een boek gegeven.
'John has given Pete a book'
(2) Jan heeft een boek aan Piet gegeven.
'John has given a book to Pete'

In contrast to and as an elaboration on such approaches, this paper gives centre stage to the constructional semantics of the Dutch prepositional-dative construction with aan. The semantic range of this construction will be examined on the basis of real language data from a 9,5 million word corpus of contemporary Dutch newspaper language. Unsurprisingly, the prepositional-dative construction is frequently attested with prototypical verbs of giving such as geven 'give' and overhandigen 'hand', verbs of communication such as vertellen 'tell' and schrijven 'write', verbs of future transfer such as beloven 'promise' and aanbieden 'offer', etc., i.e. all kinds of verbs which are frequently attested in the ditransitive construction as well. However, next to these alternating verbs, the prepositional-dative is often used with a variety of non-alternating verbs, too, i.e. verbs which cannot enter in the ditransitive construction and which are subsequently hardly ever taken into account in existing studies. Good examples from various semantic classes are ontlenen 'derive', verliezen 'lose', toevoegen 'add' and besteden 'spend, devote', all of which turn out to be significantly attracted to the prepositional-dative construction according to the "collexeme analysis" method introduced by Stefanowitsch & Gries (2003).

The aim of this paper is to provide a corpus-based overview of the semantic structure of the Dutch prepositional-dative construction, following the multidimensional approach to constructional semantics advocated in Geeraerts (1998). Ultimately, such an account will provide a better insight into the exact semantic relation between the prepositional-dative and the ditransitive as well. Diachronic corpus data on the semantic history of the preposition aan will be brought in to shed more light on the semantic links between the various uses of the synchronic construction. Incidentally, the differences between the Dutch construction with aan and the English construction with to will turn out to be larger than one might suspect at first sight.

Geeraerts, D. 1998. The semantic structure of the indirect object in Dutch. *The Dative II*. W. van Langendonck & W. Van Belle (eds), 185-210. Amsterdam: Benjamins.
Goldberg, A. 2002. Surface Generalizations: an alternative to alternations. *Cognitive Linguistics* 13: 327-356.
Stefanowitsch, A. & S. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8: 209-43.

A usage-based approach to constructional ambiguity: the case of Dutch participles

Gert De Sutter
University of Ghent

In the present paper, we argue that the structural properties of a construction can be affected by constructional ambiguity, i.e. the ambiguity that arises in actual language behaviour when two or more constructions are formally identical, but semantically/functionally distinct. More particularly, we show how the constructional ambiguity between Dutch embedded [PART Vfin] and embedded [ADJ Vfin] constructions (Vfin = finite verb), in which both ADJ and PART are instantiated by a participle, influence the degree of word order variability of these constructions. Traditionally, it is assumed that when the participle occurs in the [PART Vfin] construction, it inherits both the constructional properties "action" semantics and variable sequential order (the participle either precedes or follows the auxiliary verb). When the participle occurs in the [ADJ Vfin] construction, it is assumed to inherit the adjectival properties of the construction, i.e. the "state" semantics and invariable sequential order (the adjectival participle always precedes the verb).

However, some researchers have observed that [ADJ Vfin] constructions may exhibit word order variability when participles instantiate the construction. In order to test this hypothesis empirically and quantitatively, we first have to deal with the following double-sided problem: (i) both constructions are formally not distinguishable from each other and (ii) the semantics/function of the ambiguous constructions under scrutiny is hard to access (Akerman & Goldberg 1996, Levin, B. & M. Rappaport(1986). Hence, the methodological question to be answered first is:

How do we discriminate on a concrete level (e.g., in corpus data) which construction is activated? Formulated somewhat differently, how do we discriminate between the adjectival and the verbal use of participles, in such a way that the word order effect can be tested in a quantitative usage-based study?

The present paper presents a procedure which enables a fine-grained, non-intuitive, systematic and reliable detection of the actual status of a participle, and, hence, the actual status of the construction. The procedure takes into account (i) contextual cues, (ii) probabilistic information (derived from Celex and the CGN corpus of spoken Dutch, and (iii) distributional properties of the participles. The output of the procedure is both binary (adjectival participle vs. verbal participle) and continuous (on a scale from extremely verbal to extremely adjectival). The latter type of output meets the objections of Cognitive Linguistics and other usage-based frameworks, by treating the category boundary between adjectival and verbal participles not as clear-cut, but as gradual. Moreover, a continuous classification also allows for a more fine-grained analysis and more powerful statistical techniques.

The results show that some participles are more eligible to be incorporated in the [ADJ Vfin] construction than other participles, i.e. participles exhibit varying degrees of adjectivity/verbality. Moreover, the hypothesis that the [ADJ Vfin] construction exhibits some word order variability is confirmed, even though there is a very strong statistical correlation between the use of the [ADJ Vfin] construction and the absence of word order variability. The consequences of these results for Construction Grammar will be discussed.

Ackerman, F. & A. Goldberg (1996). Constraints on Adjectival Past Participles. *Conceptual structure, discourse and language*. A. Goldberg (ed.), 19-30. Stanford: CSLI.

Levin, B. & M. Rappaport (1986). The formation of adjectival passives. *Linguistic inquiry* 17: 623-661.

Quantifying the interaction between modality and aspect in Slavic languages

Dagmar Divjak & Oesten Dahl
University of Sheffield and Belgian Science Foundation & Stockholms Universitet

Much effort has been put into clarifying the relation between modality and other verbal properties, in particular mood and tense. Until recently (cf. the workshop by Abraham & Leiss at SLE 2006) the relation between modality and aspect received much less attention, however. For Slavic languages this situation is particularly unfortunate: Slavic deviates from the cross-linguistic aspectual norm by marking the imperfective versus perfective distinction on all verbal forms (Dahl 1985: 74-85). Moreover, the hypothesis that directed much of the research, i.e. that imperfective aspect prevails in modal constructions or that the imperfective is used to express epistemic or alethic modality whereas perfective aspect renders deontic meanings, seems too simplistic for Slavic data.

We present a corpus-based, quantitative and qualitative analysis of the interaction between aspect and modality in an East, West and South Slavic language, i.e. Russian, Polish and Serbian. The constructions we focus on are positive statements built around a modal word followed by an infinitive, e.g., *Zdes' možno perechodit'impf ulicu* ('you can cross [permissibility] the street here') vs *Zdes' možno perejtipf ulicu* ('you can cross [possibility] the street here'). Starting point for the comparative study is the situation in Russian that provides *moč'* and *možno* to express ability, possibility and permissibility but *dolžen*, *nužno*, *nado*, *neobxodimo*, *sleduet* and *prixoditsja/prišlos'* to express necessity and obligation. On the basis of data extracted from a small parallel Slavic corpus compiled specifically for this study, Polish and Serbian translational equivalents have been identified to facilitate a direct comparison with the findings for Russian.

Corpus data reveal that the perfective infinitive is used three times more frequently in constructions with modal words in all three languages. Therefore, we have used the perfective as starting point: having determined the range of situations a perfective infinitive can capture, we will turn to investigating which properties (dis)appear when the imperfective is used (cf. Forsyth 1970: 263-271). In order to reveal significant differences between modal constructions with imperfective and perfective infinitives the dataset is being tagged, in first instance, for properties that relate to the semantics of the modal word (the modality type expressed, e.g., epistemic vs deontic, possibility vs permissibility vs ability etc) and of the aspect of the infinitive (the aspectual type rendered, e.g., generalizing vs specifying, activity focused vs action/result focused) as well as to the degree of control (high, medium, low) the subject has over the infinitive action and the level of intentionality (intentional vs accidental) with which s/he carries out that action. In second instance, lexical semantic information will be taken into account such as the semantic type the infinitive action belongs to or the adverbial specification found in the sentence (cf. the scheme proposed in Divjak & Gries 2006).

Statistical analysis of the data, viz. logistic regression, will be used to reveal the variable or set of variables that has the highest predictive power for aspect assignment in a particular type of modal construction. These findings will shed light on the complex interaction of a multitude of variables that drive aspectual variation in modal constructions.

Dahl, Ö. 1985. *Tense and Aspect Systems*. Oxford: Blackwell.

Divjak, D. & St. Th. Gries, 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2: 23-60

Forsyth, J. 1970. *A Grammar of Aspect. Usage and Meaning in the Russian Verb*. Cambridge: Cambridge University Press.

Comparable Corpus Elicitation for Cognitive Semantic Analysis

Kerstin Fischer
University of Hamburg

Many linguistic forms are part of systems of linguistic choices that differ only in subtle cognitive-functional respects which often concern their appropriateness in, and association with, particular situations. While such relationships have traditionally been described in terms of register, in a usage-based model such aspects have to be represented as part of their cognitive semantic description, and cognitive linguistic formalisms, like Embodied Construction Grammar (ECG), have been particularly designed to account for such relationships.

An aspect that further complicates the matter is that situation is not objectively given (Lakoff 1987). Correspondingly, empirical studies of linguistic use in objectively identical situations have revealed both inter- and intrapersonal variation (Fischer 2006a). Halliday and Matthiessen (2004) thus hold registers to be probabilistic. However, in a cognitive semantic analysis the aim is to identify all aspects that explain the meaning and usage of a given form, and thus it cannot stop at the point of simply recording probabilities. Accordingly, it can be shown that much situational variation can be explained by the fact that speakers construe what the situation consists in differently; the key problem then becomes to identify the speakers' own cognitive models of the situation and to relate the occurrence of linguistic features to these models (Fischer 2006b) in order to tease out the subtle cognitive-functional aspects that distinguish the usage of the linguistic forms of a linguistic subsystem.

The procedure that I propose here is to elicit comparable corpora that differ only in single situational variables and to analyse these corpora both qualitatively and quantitatively. Qualitative analyses, based on principles of ethnomethodological conversation analysis, allow us to identify the speakers' situation models as participant categories, that is, as the participants' own models, quantitative analyses then reveal statistical correlations between features indicating different situation models and the usage of the linguistic construction under consideration.

To illustrate the method proposed, I present a detailed study of grammatical mood in corpora of human-robot interaction that (objectively) differ only with respect to a single feature (set I differs with respect to the appearance of the robot; in set II the same robot produces differing verbal output). In the different corpora, mood choice can be shown to be significantly related to the speakers' different concepts of the human-robot interaction situation and to other linguistic features indicating differing cognitive representations of the artificial communication partner. I then propose a cognitive semantic analysis of grammatical mood, formalizing the interaction between the speakers' situation models and the grammatical choice in ECG (Chang et al. 2002).

Chang, N., Feldman, J., Porzel, R. & Sanders, K. 2002. Scaling Cognitive Linguistics: Formalisms for Language Understanding. *Proceedings of the 1st International Workshop on Scalable Natural Language Understanding*, Heidelberg 2002.

Fischer, K. 2006a. *What Computer Talk Is and Isn't: Human-Computer Conversation as Intercultural Communication*. Saarbrücken: AQ.

Fischer, K. 2006b. The Role of Users' Preconceptions in Talking to Computers and Robots. *Proceedings of the Workshop on How People Talk to Computers, Robots, and other Artificial Communication Partners*, Delmenhorst, April 21-23, 2006. SFB/TR8 Report No. 010-09/2006.

Halliday, M.A.K. & Matthiessen, C. 2004. *Introduction to Functional Grammar*. London: Arnold.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

The cognitive reality of frequent verb-noun combinations: An empirical study

Gaëtanelle Gilquin
FNRS, Université catholique de Louvain

The cognitive reality of frequency is often taken for granted by linguists. If a particular linguistic phenomenon is frequent, the argument goes, it must have some relevance for cognition. Several studies have indeed suggested that frequency of occurrence has a place in the cognitive system and that information about frequency might actually be stored in memory by an automatic encoding process (cf. Hasher 1984). This would lead to the expectation that frequent items are somehow more salient or, as Schmid (2000: 39) puts it in the "From-Corpus-to-Cognition Principle", that "[f]requency in text instantiates entrenchment in the cognitive system".

Some recent studies, however, have emphasised the discrepancy that may exist between frequently attested items and cognitively salient ones (e.g. Shortall, in preparation). In an earlier study (Gilquin 2005, 2006), I showed that the most frequent sense of a word does not necessarily correspond to the sense that comes first to mind when prompted for this word. More particularly, I demonstrated that, while highly polysemous verbs such as take or give are mostly used in their delexical sense in a corpus of spontaneous conversations (e.g. take a walk or give a smile), people seem to access more concrete senses first (e.g. take somebody home or give somebody a book).

In this presentation, the focus will be on combinations of a high-frequency verb and a noun phrase, and the approach will be more lexical. Applying Gries & Stefanowitsch's (2004) method of distinctive collexeme analysis to the comparison of one construction in two language varieties (rather than two constructions in one language variety), I will compare the noun phrase of such combinations in free language production (corpus data) and in controlled language production (elicitation data). The degree of overlap and the rate of lexical variation will be measured, and the grouping of distinctive collexemes into semantic fields will be examined. The observation of the differences and similarities between the collostructions in corpus and elicitation data will provide greater insight into the question of the cognitive reality of linguistic frequency.

Gilquin, G. 2005. What You Think Ain't What You Get: Highly polysemous verbs in grammar and mind. Paper presented at the conference From Gram to Mind: Grammar as Cognition, Bordeaux, 19-21 May 2005.

Gilquin, G. 2006. Language production: a window to the mind?. Paper presented at the 22nd Scandinavian Conference of Linguistics, Aalborg, 19-22 June 2006.

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

Hasher, L. 1984. Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist* 39 (12): 1372-1388.

Schmid, H.-J. 2000. *English abstract nouns as conceptual shells. From corpus to cognition*. Berlin: Mouton de Gruyter.

Shortall, T. In preparation. Corpus, Cognition, and Curriculum. PhD Thesis, The University of Birmingham, Birmingham.

Testing the hypothesis. Confirmatory statistical techniques for multifactorial data in Cognitive Semantics

Dylan Glynn, Dirk Geeraerts, & Dirk Speelman
University of Leuven

This study examines quantitative techniques for the analysis of polysemy and parasynonymy. As Cognitive Linguistics continues to embrace corpus-driven methodologies, we are faced with a growing need to implement quantitative methods of data analysis (Tummers & al. 2005, Heylen & al. in press). Current trends in the study of polysemy have focused on exploratory techniques such as Cluster Analysis (Gries 2006) and Correspondence Analysis (Glynn in press). The goal of such exploratory statistics is to identify and visualize patterns in the data. These patterns are argued to represent patterns of usage, and therefore polysemy.

The importance of these techniques notwithstanding, the cognitive framework needs to deepen its use of quantitative research especially through the use of confirmatory multivariate statistics. Exploratory statistic analysis does not permit inferences about the language, only the sample, or dataset, investigated. However, in confirmatory statistics, inference is made from the sample to the population. In other words, one claims that what one sees in the data is representative of the language generally.

There exists a range of models for multivariate statistical analysis some of which are gaining currency in especially constructional parasynonymy. Within Cognitive Linguistics for instance, Linear Discriminant Analysis (Gries 2003, Wulff 2004) and Logistic Regression Analysis (Heylen 2005, De Sutter & al in press) represent two such techniques. These have been successfully used to capture the various conceptual, formal, and extralinguistic factors that lead to the use of one construction over a another. However, the study of polysemy differs at this point. Instead of examining the variables that effect the use of one parasynonymous forms to another, we are examining the interaction of a range of formal variables (the lemma and its syntagmatic and inflectional variation), semantic variables, and extralinguistic variables, in the search for correlations across all of these. One possible multivariate technique for this type of data is Log-Linear Modelling.

We apply Log-Linear Analysis to the semantic and formal variation of the parasynonymous lemmas bother, annoy, and hassle in British and American English. Using a large non-commercial corpus, 500 verbal instances are annotated for semantic, formal, regional features. The semantic variables include argument types, argument relations, and case roles. The formal annotation includes tense, aspect, and argument structure. Relative to the two dialects in question, a complex set of correlations between the different formal and semantic variables are identified through exploratory studies. These correlations are then modelled using Log-Linear Analysis. The results are interpreted and the efficacy of the method is critically assessed.

De Sutter, G. & al. In press. Detecting and balancing determinants of word order variation in Dutch clause final verb clusters. *Linguistics*.

Glynn, D. in press. Polysemy, syntax, and variation. A usage-based method for Cognitive Semantics. *New Directions in Cognitive Linguistics*. V. Evans & S. Pourcel (eds). Amsterdam: Benjamins.

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

- 2006. Corpus-based methods and Cognitive Semantics: The many senses of to run. *Corpora in Cognitive Linguistics*. St. Th. Gries & A. Stefanowitsch (eds), 57-100. Berlin: Mouton.

Heylen, K. 2005. A Quantitative Corpus Study of German Word Order Variation. *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. S. Kepser & M. Reis (eds), 241-264. Berlin: Mouton.

- Tummers, J., & Geeraerts, D. In press. Methodological issues in corpus-based Cognitive Linguistics. *Cognitive Sociolinguistics*. G. Kristiansen & R. Dirven (eds). Mouton: Berlin.

Tummers, J., Heylen, K., & Geeraerts, D. 2005. Usage-based approaches in Cognitive Linguistics. State of the Art. *Corpus Linguistics & Linguistic Theory* 1: 225-261.

Wulff, S. 2003. A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8: 245-282.

Quantifying meaning: where do the English and Russians start/begin?

Stefan Th. Gries & Dagmar Divjak
University of California, Santa Barbara & University of Sheffield.

Ever since the emergence of Cognitive Linguistics as a research paradigm, the analysis of semantic structures has been high up on the cognitive-linguistic agenda. Early studies, which shaped the field for years to come, investigated the degree to which conceptual mappings (metaphor and metonymy) as well as polysemous radial categories allowed for new insights into the linguistic organization – if not also mental representation – of semantic knowledge.

In spite of the wealth of new insights yielded by these works, recent years have witnessed a growing recognition of shortcomings related to one of the core commitments underlying cognitive-linguistic research: usage is correlated with grammatical knowledge and with the mental representation of that knowledge. In the 1990s the field saw a lively discussion that was concerned with the degrees to which (i) distinctions arrived at by linguists reflect naïve speakers' intuitions and (ii) linguistic data alone license claims about mental representations. In addition, in spite of the popularity of the term "usage-based approach", the number of works using actual corpus-data has – until recently – been fairly small.

In the present study, we will be concerned with the theoretical and methodological issues involved in building cognitive linguistic theory on actual usage data. In earlier work (2006, to appear) we have established the notion of a "behavioral profile" for an extremely fine-grained analysis of the distributional behavior of nine near synonymous verbs in Russian and a highly polysemous verb in English. This approach involves the largely manual analysis of many different distributional characteristics of the search words, including morphological characteristics (tense, mode, aspect, voice etc.), syntactic characteristics (sentence and clause type) as well as semantic characteristics (for things this includes reference to their thematic roles, countability etc., for processes this captures controllability, the properties of the process they denote, etc.).

Here, we will adopt and refine our methodology to explore the distributional behavior of two near synonymous verbs in English (*begin* vs. *start*) and Russian (*načat'* vs *stat'*). Based on data from the Uppsala corpus of Russian and the British Component of the International Corpus of English, we will show that the behavioral profile approach, more than comparable approaches, 1) yields data that facilitate providing a fine-grained and bottom-up analysis of differences between semantically closely-related words within one language as well as contrasting data on semantically similar words from two languages; 2) makes it possible to study issues that have long occupied cognitive linguists, e.g. the semantic import of different complementation patterns such as *-ing* forms vs. infinitives in English and how these semantic differences are rendered in Russian, a language that only has infinitives at its disposal.

Divjak, D. & Gries, St. Th. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2.1:23-60.

Gries, St. Th. & Divjak, D. to appear. Behavioral Profiles: a corpus-based approach to cognitive semantic analysis. *New Directions in Cognitive Linguistics*. V. Evens & S. Poucel (eds). Amsterdam: John Benjamins.

Analyzing constructional pragmatics:
a usage-based study of German word order constructions.

Kris Heylen
University of Leuven

The verb arguments subject, dative object and/or accusative object can occur in different orderings in the so-called Middle Field of the German clause. Each ordering has subtly different syntactic, semantic and pragmatic properties, which poses both a theoretical and methodological challenge. Theoretically, the word order variants can be regarded as alternating, highly schematic construction schemas. Their modelling requires the integration of a set of heterogeneous factors into each construction's description, an explicit characterization of the relation between the alternating constructions, as well as a characterization of their relation to other (more specific) constructions. Methodologically, advanced corpus-based analysis is necessary to untangle the complex interplay of the many factors. In this study, we look at a minimal ordering pair from the point of view of a usage-based Construction Grammar. We will use two types of statistical corpus analysis to determine the different aspects of the usage characteristics of the two word order variants. Their main functional difference is the way they allow the speaker to construe the perceived prominence of the referents. We will therefore interpret the word order variants as constructions with specific pragmatic properties. The data also shows that the alternation fulfils this function in a range of more specific contexts, which suggests that the two constructions should at least partially be described at a lower level of schematicity.

The specific word order variants we will look at occur when both a nominally realized subject and a pronominally realized object are present in the Middle Field. In that case the pronominal object can either precede the subject (a) or follow it (b), without any noticeable difference in meaning or grammaticality.

(a) Dabei hat ihm der Bruder geholfen (his brother has helped him with that)
(b) Dabei hat der Bruder ihm geholfen

In a first corpus analysis, all instances (n=995) of the two orderings in the NEGRA newspaper corpus (350 K tokens) are analyzed in a multivariate statistical model to investigate the effect of some 20 explanatory variables on the choice of word order. These include morpho-syntactic, semantic and pragmatic variables. The analysis indicates that one information-structural factor has an especially strong effect, viz. the presence of a competing subject referent in the previous context. This identifies the marked subject-first order (at least partially) as a pragmatic construction. Compared to the default constructional option, the marked order gives prominence to the subject's referent in order to contrast it with alternative referents.

A second corpus analysis zooms in on the specific contexts in which the subject-first order realizes its pragmatic function. With a distinctive collexeme analysis (Gries & Stefanowitsch 2004), we investigate the typical (i.e. statistical significant) lexical instantiations of the clause slots that co-occur with the marked subject first order. To avoid data sparseness, the much larger Tübingen Partially Parsed Corpus (200M tokens) was analyzed semi-automatically. The analysis indicates a.o. that the marked order occurs typically after concessive complementizers or with verbs of conflict, which both prime a contrast between referents in discourse. This confirms the results of the first corpus study and indicates that marked order construction is closely linked to more specific (pragmatic) constructions within a hierarchical network of constructions.

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

On the futurate present in English and German

Martin Hilpert
Rice University

This paper argues for a constructional account of the so-called futurate present in English and German. In both languages, sentences in the present tense can make reference to the future, as illustrated below.

(1) a. The Yankees play the Red Sox tomorrow. (Lakoff 1971: 339)
b. Der HSV spielt morgen gegen Frankfurt.

The present tense thus functions as an alternative to other available future constructions. While forms such as English *be going to* or German *werden* are generally recognized as future constructions, it is not self-evident that the respective futurate presents should be analyzed as symbolic units that conventionally denote future time. After all, such an account contradicts the intuitively plausible notion that present tense forms simply refer to the present, and only deviate from this temporal reference in the case of contextual override (d'Alquen 1997). This paper uses quantiative evidence to weigh these competing hypotheses.

This paper uses the method of distinctive collexeme analysis (Gries and Stefanowitsch 2004) in a way that diverges to some extent from previous applications. The usual starting point of a collostructional method is a corpus search that exhaustively retrieves all tokens that instantiate a given construction. The present analysis departs from this guideline and starts with an exhaustive retrieval of future time adverbials such as English *tomorrow* or German *nächste Woche* from large balanced corpora.

A concordance of the word *tomorrow* will contain sentences with *be going to*, *will*, *shall*, and crucially, present tense forms that convey future meaning. For the purposes of the present study, the retrieved examples are divided into three categories. Examples (2a) to (2c) below represent the futurate present, other future constructions, and examples without future time reference.

(2) a. So it's your birthday tomorrow, said Lucy. (a-c: BNC)
b. I'll call you tomorrow, said Lucy.
c. Where today's talent is turned into tomorrow's champion.

In both future-denoting categories, the main verbs of the examples are identified and lemmatized, such that example (2a) represents a token of *be*, and (2b) is counted as a token of *call*. This procedure yields two raw frequency lists that can serve as the input for a distinctive collexeme analysis. For both English and German, the method identifies verbs that are significantly uneven in their distribution. These lexical preferences reflect the respective constructional meanings of the futurate present in English and German, revealing differences between the two that pertain to subject animacy, agentivity, and lexical aspect. It is concluded that these usage patterns motivate a recognition of the futurate present in the investigated languages as constructions (Goldberg 2006).

D'Alquen, R. 1997. *Time, Mood and Aspect in German Tense*. Frankfurt: Peter Lang.
Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
Gries, St. Th. & Stefanowitsch, A. 2004. Extending collostructional analysis: A corpus-based perspectives on 'alternations'. *International Journal of Corpus Linguistics* 9: 97-129.
Lakoff, G. 1971. Presupposition and relative well-formedness. *Semantics*. D. Steinberg & L. A. Jakobovits (eds), 329-40. Cambridge: Cambridge University Press.

A collostructional analysis of the causative alternation in English

Maarten Lemmens
Université Lille3 & CNRS

The present paper discusses the interaction of verbal and constructional semantics as apparent in the causative alternation with English lexical causatives, e.g., He bounced the ball vs. The ball bounced, also called the unaccusative alternation.

Most of the (generative inspired) work on unaccusativity (e.g., Levin 1993, Levin & Rappaport 1995) posits some form of "derivation principle", where different structures (in our case, the non-causative) are said to be derived from one and the same deep structure (in our case, the causative). Cognitive Linguistics, in contrast, takes a monostratal view and emphasizes the need to consider each construction in its own right. Recent publications (e.g., Goldberg 2002, Yoshimura & Taylor 2004) insist quite strongly on "surface generalisations" as an alternative to an approach based on alternations.

The study reported on in this paper aligns itself fully with the cognitive monostratal view on constructions, and even goes one step further, by looking in more detail at the particular contexts in which the causative alternation can arise. We present results from our corpus-based analysis of causative and non-causative constructions for a selection of English lexical causatives, taken from different kinds of verb classes, such as BREAK verbs (break, burst, crumble, etc.), MOTION verbs (roll, bounce, etc.) or verbs from other domains (e.g., expand, worsen). More specifically, our study, based on data from the British National Corpus, is a type of distinctive colloxeme analysis (cf. Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2004) studying the variation pertaining to one semantic role slot in related constructions. The slot in question in our study is that of the Theme affected by the process (e.g., ball in the example above). Our study reveals that the overlay of this participant is surprisingly low for the different constructions, which lends support to the idea that the semantics of the construction selects its 'own' type of thematic participant. The present study also confirms the results of earlier collostructional analyses that the alignment of specific lexemes and constructions is much stronger than is often assumed in the literature.

Our usage-based study can thus also be taken in support of a "surface generalisation" approach, going even further than individual verbs by looking at their specific uses as well. At the same time, we show that much larger generalisations, which underlie the general alternations that have been observed in the literature, are still pertinent to account for certain phenomena (e.g., constraints, prototype effects, etc.) that otherwise remain inexplicable. As such, we provide some further insights into the lexical and constructional interface in the English grammar.

Goldberg, A. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13: 327–356.

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

Levin, B. 1993. *English Verb Classes and Alternations. A Preliminary Investigation*. Chicago: Chicago University Press.

- & Rappaport Hovav, M. 1995. *Unaccusativity. At the Syntax-Lexical Semantics Interface*. Cambridge, Mass.: MIT.

Stefanowitsch, A. & S.Th. Gries. 2003. Collostructions: investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8: 209-243.

Yoshimura K. & Taylor, J. 2004. What makes a good middle? The role of qualia in the interpretation and acceptability of middle expressions in English. *English Language and Linguistics* 8: 293-321.

The Semantics of Language Use: Why Corpus Matters

François Nemo
University of Orleans

The aim of my presentation will be to show that when it comes to understanding the cognitive mechanisms central to language, language use and semantics, a data-driven methodology can provide new insights for cognitive linguistics and cognitive semantics.

In regard to data, linguists may indeed adopt three approaches:

- work on chosen fragments ("central" uses), adopting explicitly or implicitly the idea that semantic meaning is accessible through intuition and that intuition is methodologically acceptable;
- take into account a wide range of uses and meanings, but postulate a hierarchy between them and distinct mechanisms to produce "central" uses and "non central" uses;
- consider all uses to be equal and the diversity of uses as an indication that the mechanisms at stake are different from what intuition first indicates.

In the first cases, the methodological risks are circularity, self indoctrination and unfalsifiability. This is because either one chooses to ignore or marginalize some uses, which is a theory in itself and one that cannot be tested, or because if polyfunctionality and polysemy are the rule and not the exception, the distinction between central and non central uses becomes a protective belt, in the Lakatosian sense, preventing the falsification of the "central" mechanisms.

What I shall show through the study of a variety of linguistic items, from discourse particles (Nemo 2006, 2007) to nouns (Nemo 2002) and affixes, is that adopting the third approach and using a Data-Driven methodology i) raises important questions; ii) allows the formulation of a morpheme/construction framework and methodology; iii) contradicts, to a certain degree, the Cognitive Commitment advocated by Lakoff (1990) and Evans & al. (in press). This commitment holds that "quite a lot is known about human categorization, […] a theory that reduces word meaning to the same mechanisms responsible for categorization in other cognitive domains is simpler than one that hypothesizes a separate system for capturing lexical semantics" (Evans & al.in press).

To support the above claim, the study demonstrates that in order to account for the puzzling distribution and uses of linguistic items, we need may need integrate other cognitive mechanisms that have been largely ignored in cognitive research. This entails that the exact role of cognitive mechanisms, often assumed to be central in language and cognition(Langacker 1991), may have to be reconsidered.

Evans, V, Bergen B. & Zinken, J. in press. The Cognitive Linguistics Enterprise: An Overview. *The Cognitive Linguistics Reader*. London: Equinox.

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

Lakoff, G. 1990. The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics* 1: 39-74.

Langaker, R. 1990. *Concept, Image, Symbol: The Cognitive Basis of Grammar*. Berlin: Mouton de Gruyter.

Nemo, F. 2002. Indexicalité ou catégorisation : le sens entre signification et dénomination. *La représentation du sens en linguistique*. P. Larrivée & D. Lagorgette (eds), 49-68. Munich: Lincom Europa.

Nemo, F. 2006. Discourse words as morphemes and as constructions. *Approaches to Discourse Particles*. K. Fischer (ed.), 415-449. Amsterdam: Elsevier.

Cuyckens, H., Dirven, R. & Taylor. J (eds). 2003 *Cognitive Approaches to Lexical Semantics*. Berlin: Mouton de Gruyter.

Balancing Acts: Empirical Pursuits in Cognitive Linguistics

John Newman
University of Alberta

Cognitive Linguistics, as practised in the International Cognitive Linguistics Conferences, has not always been particularly empirical in its orientation. This is changing, however, as Cognitive Linguistics is increasingly incorporating a range of empirical methodologies into its research activities, reflecting a larger trend within the field of Linguistics. This development is welcome, correcting, as it does, the over-reliance on relatively artificial, decontextualized data. The idea of a "usage-based " approach is symptomatic of this trend. In this talk, I review multiple ways in which one might implement such an approach. As alternative foci for usage-based research, for example, one might investigate an individual's language usage vs. societal language usage, variation in usage (genre, age, gender etc.), sentences vs. utterances, usage as reflected in typical transcripts vs. usage as observed in the communicative acts, production vs. reception, language-specific phenomena vs. cross-linguistic phenomena. As alternative empirical methodologies, one could adopt experimental techniques, corpus-based methodologies, or other techniques which exploit large databases.

The presence of so many competing agendas, all claiming for attention and validation, is, one would like to think, healthy for the discipline. It means, though, that there must be a constant "balancing" of approaches against each other if we are to achieve a fully inclusive and informed approach to the study of language. Common to most of these alternative agendas is an interest in data, ideally in large amounts, leading inevitably to questions about the statistical validity of methodologies and significance of results. Within the field of corpus linguistics, in particular, where we are dealing with connected discourse, there are unresolved issues concerning the most appropriate statistical methods to use (see, for example, Dunning 1993, Kilgarriff 2005, and Gries 2005). I consider a variety of approaches to "quantification" and argue that the interests of Cognitive Linguistics are best served by a tolerance for methodical (but statistically unproven) procedures for some kinds of data, alongside statistically rigorous procedures where these are appropriate (cf. McEnery and Wilson 2001: 75-77).

Dunning, T. E. 1993. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics* 19: 61-74

Gries, St. Th. 2005. Null-hypothesis significance testing of word frequencies: A follow-up on Kilgarriff. *Corpus Linguistics and Linguistic Theory*. 1.2: 277-294.

Kilgarriff, A. 2005. Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory*. 1: 263-276.

McEnery, T. & Wilson, A. 2001. *Corpus Linguistics: An Introduction*. 2nd ed. Edinburgh: Edinburgh University Press.

Does frequency in text really instantiate entrenchment in the cognitive system?

Hans-Jörg Schmid
Ludwig-Maximilians-Universität München

One of the basic assumptions of quantitative approaches in cognitive linguistics is that the frequencies of linguistic items or constructions in large data-bases can be related to hypothesized properties of the cognitive system. In Schmid (2000) a rather strong version of this claim was encapsulated in the "From-Corpus-to-Cognition Principle", which predicts that "Frequency in text instantiates entrenchment in the cognitive system" (2000: 39).

Four closely interrelated constructions were investigated to support this claim, using a 225-million-word corpus extracted from the *Bank of English* as a data-base:

(1) abstract N + *that*-clause: *the fact that abstract nouns are difficult to pin down ...*
(2) abstract N + *to*-infinitive: *the idea to illustrate the patterns investigated ...*
(3) abstract N + BE + *that*-clause: *the problem is that there is a lot to study*
(4) abstract N + BE + *to*-infinitive: *the solution is to focus on a bunch of examples*

From a methodological point of view the advantage of this set of constructions is that there is a clearly definable relation between grammatical patterns and a limited number of lexical items filling the nominal slot in the construction. As a result, fairly simple mathematical measures could be used to assess the relative frequencies of types of nouns in the patterns (vis-à-vis their overall frequencies) and the reciprocal dependence of the lexical and the grammatical components of the construction. These measures were dubbed *attraction* (the degree to which a grammatical pattern attracts a particular noun) and *reliance* (degree to which a noun relies on a grammatical pattern for its occurrence). Mathematically simple-minded and restricted in application to the patterns investigated, the two measures are precursors to more sophisticated work in the collostructional framework proposed later by Stefanowitsch and Gries (2003; Gries and Stefanowitsch 2004).

In my paper I want to review the main methodological and theoretical assumptions of Schmid (2000) and relate them to later work in the collostructional paradigm. In hindsight, a critical view of the methodological assumptions and the results of the study is also in order, however. While it still seems plausible that high-frequency nouns with high attraction scores serve as cognitive anchors for certain patterns (e.g. *the fact that*-clause or *the aim is to*-infinitive), the roles played by medium-frequency nouns and by rare nouns with very high reliance scores (e.g. *the temerity to …*, *the realisation that …*) is much harder to assess and throws the general applicability of the From-Corpus-to-Cognition Principle, as well as a too blind reliance on frequency data in general, into doubt.

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

Schmid, H.-J. 2000. *English abstract nouns as conceptual shells. From corpus to cognition*. Berlin: Mouton de Gruyter.

Stefanowitsch, A. & Gries, St. Th. 2003. Collostructions: investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8: 209-243.

Information structure and constructional semantics

Anatol Stefanowitsch
University of Bremen

It has been proposed that information structure interacts with argument structure in crosslinguistically observable systematic ways such that core argument positions differ in the ease with which they accept noun-phrases with discourse-new referents. Specifically, Du Bois (1987) shows that in ergative-absolutive languages S (the logical subject of intransitive clauses) and O (the logical object of transitive clauses) readily accept both discourse-new and discourse-old referents, while A (the logical subject of transitive clauses) accepts only discourse-old referents. He suggests that this relationship between grammatical relations and information structure may be universally valid and argues that it provides a direct motivation for the existence of ergative alignment systems in terms of information processing constraints.

Studies of nominative-accusative languages like English consistently replicate DuBois' results with respect to A and O, but they yield less clear results with respect to S. S typically falls between A and O with respect to its information structural properties, but is generally closer to A (Kumagai 2001, Kärkkainen 1996, Sakita 2005, Stefanowitsch 2006).

In my paper, I show that the relationship between grammatical relations and information structure is considerably more complex than suggested by DuBois. I use a large representative sample of English intransitive and transitive clauses manually annotated for the information status, lexical category and animacy of S, A and O as well as the semantic category of the verb and the construction type of the clause. I apply a multivariate statistical method called hierarchical configural frequency analysis (HCFA, cf. von Eye 2002) to this data set to identify configurations of linguistic features that can be considered entrenched schemas in the sense of Langacker (1987).

A careful analysis of these schemas shows that the category S is too broad to display uniform information-structural properties, and that regularities in the relationship between information structure and argument structure in English are more accurately captured at the level of construction-specific roles. Moreover, these regularities largely follow from semantic properties of the constructions in question.

Du Bois, J. 1987. The discourse basis of ergativity. *Language* 63: 805-855.

Kumagai, Y. 2001. Preferred Argument Structure and discourse ergativity in English. *Mulberry* 50: 77-90.

Kärkkäinen, E. 1996. Preferred Argument Structure and subject role in American English conversational discourse. *Journal of Pragmatics* 25: 675-701.

Sakita, T. 2005. Discourse Basis of Grammatical Patterns: Preferred Argument Structure in Narrative. Presented at the 9th International Pragmatics Conference, Riva del Garda, 10-15 July 2005.

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

Stefanowitsch, A. 2006. Information structure and argument structure: The case of S. Paper presented at the 2nd DGKL Conference, Munich.

von Eye, A. 2002. *Configural Frequency Analysis: Methods, Models, and Applications*. Mawah: Lawrence Erlbaum.

Marrying cognitive-linguistic theory and corpus-based methods:
On the compositionality of English V-NP idioms

Stefanie Wulff
University of Bremen

Non-compositionality, the degree to which a phrase's meaning is a function of the meanings of its parts, continues to play a key role in constructionist approaches to grammar (Goldberg 2006). This paper presents a corpus-linguistic model of compositionality that is based on work by Berry-Rogghe (1974), who measured compositionality by determining the collocational overlap between the head word of the phrase and the phrase itself. On the basis of more than 13,000 tokens of V NP-patterns like *take the plunge or have a laugh* extracted from the *BNC*, this paper describes an elaborated compositionality measure that implements some of the most fundamental assumptions of cognitive-linguistic approaches:

- Not only one, but all component words of the phrase are taken into account. That is, rather than considering, say, only the verb *make* in *make a point*, the measure also quantifies the contribution made by *point*. Therefore, the measure accords with the constructionist view that a complex phrase is a manifestation of several smaller constructions, with every one of them contributing to the meaning of the complex construction (Goldberg 2006).
- Rather than assuming that each word contributes equally much to the meaning of the phrase, this compositionality measure weighs the contributions of the component words relative to each other. That is, the measure licenses the possibility that the contribution made by *point* in *make a point* can be smaller or higher than in *see a point*. Thereby, the measure implements the central assumption of cognitive approaches that constructions are differently entrenched in the mental lexicon, depending on their frequency of use (Langacker 1987); accordingly, differences in entrenchment between *make*, *see*, and *point* are expected to influence their relative weight.
- The contribution of each component word is not only weighted in terms of how much of *the pattern's* meaning is accounted for by the component word, but also how much *of itself* it brings in. For instance, in *take the plunge*, *plunge* brings in nearly all of its semantics (it shares most of its collocates with the pattern), whereas *take* only contributes a fraction of its meaning potential (most collocates of take are not associated with *take the plunge* in particular). Accordingly, the measure allows for a potential backward influence of the phrase's semantics on the (weightings in the often polysemous network of) the constituent word's semantics (Langacker 1987).

The results are shown to tie in well with established findings from psycho-linguistic research and phraseology. Ultimately, the corpus-linguistic model strengthens the case for usage-based approaches to grammar, demonstrating that a seemingly intuition-based, complex phenomenon can be modeled with performance data in a bottom-up fashion.

Berry-Rogghe, G.L.M. 1974. Automatic identification of phrasal verbs. *Computers in the humanities*. J.L. Mitchell (ed.), 16-26. Edinburgh: Edinburgh University Press.
Goldberg, A. 2006. *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.
Langacker, R. 1987. *Foundations of Cognitive Grammar.* Vol. I. *Theoretical Prerequisites*. Stanford: Stanford University Press.

Exemplars and analogy: semantic extension in constructional networks

Arne Zeschel
University of Bremen

In both Cognitive Grammar and usage-based Construction Grammar, it is customary to model constructions as complex dynamic networks that grow via piecemeal item-based extension processes. Within cognitive semantics, attempts at reconstructing such extensions have tended to focus on the role of extralinguistic (and hence, construction-independent) phenomena such as conceptual metaphor and/or certain entrenched conceptual blends, with differences in the systematicity and productivity of pertinent extensions attributed directly to these conceptual structures (e.g. Clausner and Croft 1997). By contrast, more grammatically oriented work within the usage-based paradigm often focuses on the properties of concrete linguistic elements instead (such as the type and token frequencies of different constructional types within their overall network). What is different here is that the likelihood of a certain extension is taken to depend to a large extent on statistical properties of the concrete stored exemplar(s) that it is modelled on, rather than on the properties of a certain conceptual mapping at large (e.g. Bybee 2005). Both perspectives highlight relevant aspects of analogical extension processes, and the two issues are of course connected (since what is frequent will usually be frequent for a good reason). However, they also strike a different balance between the import of genuinely constructive operations and what would have to be characterized as the more mechanistic, frequency-driven retrieval side of cognitive processing that is likewise involved in any such case of 'altered replication' (Croft 2000).

The present paper seeks to integrate these perspectives by retracing the interaction of general conceptual and item-specific frequency pressures towards/against particular semantic extensions of a constructional network. I report data from a contrastive corpus study of the lexical instantiation of three functionally equivalent constructions. Each construction comprises a number of central exemplars that have spawned more or less dense clusters of analogical variants. The study investigates the extent to which the same types and extensions from these types can be traced both across the three constructions and across the two investigated languages (thereby indicating an impact of general conceptual motivations) and, conversely, the extent to which the observed co-occurrence patterns must be characterized as individually restricted collocations. Building on work that has reported close correspondences between corpus-derived association measures and experimentally elicited acceptability judgments (Bybee and Eddington 2006) and production preferences (Gries, Hampe and Schönefeld 2005), it is argued that quantitative corpus studies hold a great potential for empirical research into the determinants of cognitive schematization processes.

Bybee, J. 2005. From usage to grammar: the mind's response to repetition. (LSA presidential address 2005)
Bybee, J. & Eddington, D. 2006. A usage-based approach to Spanish verbs of 'becoming'. *Language* 82: 323-355.
Clausner, T. & Croft, W. 1997. Productivity and Schematicity in Metaphors. *Cognitive Science* 21: 247-282.
Croft, W. 2000. *Explaining Language Change. An Evolutionary Approach*. Harlow: Longman.
Gries, St. Th., Hampe, B. & Schönefeld, D. 2005. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16: 635–676.