Handbook of Cognitive Semantics

With a Foreword by Leonard Talmy

VOLUME 2

Edited by

Fuyin Thomas Li

BRILL

LEIDEN | BOSTON

Contents

Preface VII
List of Figures and Tables VIII
Notes on Contributors XIII

VOLUME 2

PART 3
Essential Concepts

12 Figure-Ground in Cognitive Semantics 3
Rong Chen

13 Figure, Ground, and Frames of Reference in Temporal FRONT-BEHIND Relations 30
Kevin Ezra Moore

14 Closed Class vs. Open Class 77
Ye Yuan

15 Conceptualization 99
Baoyi Niu

PART 4
Semantic Categories

16 A Semantic Map for Ideophones 129
Thomas Van Hoey

17 Degree Modifiers and Scalar Meanings of Projective Grams 176
Tuomas Huumo

18 Possession in Cognitive Linguistics 205
Ricardo Maldonado

PART 5
Methodology

19 Gathering Semantic Data 237
Jürgen Bohnemeyer

20 Insights from Cognitive Semantics for Language Documentation 289
Sally Rice

21 Quantitative Corpus Methods in Cognitive Semantics/Linguistics 328
Stefan Th. Gries

22 Person-Oriented Linguistic Research 351
John Newman

PART 6
Models and Schemas

23 The Image Schema 379
Aleksander Szwedek

24 Prototypicality, Polysemy and Vagueness 409
Dirk Geeraerts

25 Cognitive Domains 433
Zeki Hamawand

Appendix: Volumes Overview 463

CHAPTER 21

Quantitative Corpus Methods in Cognitive Semantics/Linguistics

Stefan Th. Gries

1 Introduction

1.1 *General Introduction*

Quantitative methods in cognitive linguistics and in cognitive semantics have become a clear and strong "force to reckon with" (Janda, 2017: 498). Based on a quick survey of "articles proper" of the flagship journal *Cognitive Linguistics*,[1] Janda concludes that the history of the journal, and thus of the field, can be divided into two phases: (i) 1990–2007, when most articles were not quantitative, and (ii) 2008–2015, when most articles were. As argued in Sinha (2017) and esp. Janda (2017), the quantitative turn has been facilitated by a variety of developments such as

– the fact that cognitive linguistics is a usage-based theory, which in turn means it is a theory in which frequencies of exposure, use, and co-occurrence are not "just" data but in fact crucial components of "the theory/model";
– the ever greater availability of corpora and other linguistic databases;
– the ever greater availability of statistical methods.

To these, I would add the degrees to which (i) linguistics in general—not just cognitive/usage-based linguistics—has been turning more to corpus-based and statistical methods over the last 10–15 years, which exposes cognitive lin-

1 The precise definition of the papers that were included is the following: "This survey includes only articles proper, excluding review articles, book reviews, overviews, commentaries, replies, and squibs. For the purpose of this survey we define a 'quantitative article' as an article in which a researcher reports numbers for some kind of authentic language data." This definition makes me wonder (i) how articles with experimental studies were dealt with (that often involve heavy components of quantitative analysis) and (ii) how computational simulation studies would be dealt with. I think to most readers a computational small-worlds simulation with/or network-analytic statistics would definitely be a quantitative article even if it did not involve "numbers for some kind of authentic language data." It is therefore at least possible that the reported counts underestimate the number of quantitative studies in *Cognitive Linguistics*.

 | DOI:10.1163/_011

guists more to these kinds of methods and (ii) an increasing frequency of interdisciplinary collaborative work with, e.g., cognitive scientists, psycholinguists, and computer scientists/computational linguists, all fields in which quantitative analyses have been established for more longer and much more firmly than they have in cognitive linguistics.

That being said, it seems to me (very subjectively, meaning this is only an anecdotal observation for which I have no real evidence) that this development is coupled with a tendency of "the field" becoming more frequently referred to as *usage-/exemplar-based linguistics* (rather than as *cognitive linguistics*, which seems to have been the dominating term in the late 1980s and throughout the 1990s). It *also* seems to me as if, as this move towards "usage-/exemplar-based linguistics" took place, the way *usage/exemplar* was actually used especially in theoretical work developing usage-based linguistics was for the most part just frequency (absolute and relative) and association. In Section 2 of this survey, I will discuss a variety of quantitative methods—involving both observational/corpus and experimental data—that have been put to use in cognitive/usage-based linguistics and construction grammar. On the whole, that section's organization is based on the complexity and the kind of quantitative method used (to the extent that an unambiguous ranking of methods along those lines is always possible), much of the focus will be on studies that involve semantic questions such as polysemy or synonymy, but given how we cognitive linguists eschew a clear separation of the more traditional domains of syntax and lexis, several studies regarding constructional meaning will of course also be discussed. Section 3 will then briefly discuss a few recent developments and desiderata; Section 4 will conclude.

2 An Overview of Quantitative Methods and Their Applications in Cognitive Linguistics

2.1 *Monofactorial Approaches*

2.1.1 Frequencies/Probabilities

The most basic statistic we use in cognitive linguistics is frequencies of occurrences, specifically type frequencies and token frequencies. The former are concerned with, for instance, how many different linguistic elements/types are attested in a certain (constructional) slot or context. For example, Gries (2019) retrieves all instances of the so-called *as*-predicative (e.g., *Mary does not see herself as the main problem*) in the British Component of the International Corpus of English and finds 261 different verb types in the verb slot of the construction (with the usual Zipfian distribution, meaning very few types (e.g. the top 5/2%) account for a high percentage of tokens (namely 41.9%)). Type frequencies have been connected to matters of productivity, category formation, and grammaticalization (see, e.g., Bybee and Thompson 1997 or Bybee and Hopper 2001).

Token frequencies, on the other hand, are often said to be correlated with the degree to which linguistic elements might be cognitively entrenched, as is illustrated by the following famous Langacker quote (1987: 59):

> Every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence.

While statistically extremely simple, frequency—raw or transformed (e.g., the Zipf scale of van Heuven et al. 2014)—is of course one of the most widely used predictors or control variables in much psycholinguistic work, given its reliable correlation with naming/reaction times etc. To continue the above example, Gries finds an overall absolute frequency of 1131 of the *as*-predicative in the ICE-GB. However, it's probably fair to say that cognitive linguistics has relied more on relative, rather than absolute, frequencies or, put differently, on conditional probabilities: For instance, in the 1131 *as*-predicatives, the verb *see* was the most frequent verb (with 124 occurrences), which can be expressed as a conditional probability: $^{124}/_{1131} \approx 0.1096$ of the *as*-predicatives contained *see*.

One way in which conditional probabilities have proven useful is as the simplest kind of association measure: the higher the conditional probability $p(y|x)$ (read, 'probability of y given x'), the more y seems attracted to x. In one well-known study, Aslin, Saffran, and Newport (1998) show that 8-month old infants were reliably able to discriminate words and part-words (in an artificial language) based on conditional/transitional probabilities of syllable pairs, but such effects have been found on other levels of linguistic analysis as well. For example, Huang, Wible, and Ko (2012) study how differences in transitional probability make the last word of a phrase (e.g. *fact*) faster to read when it is part of a multi-word expression (e.g., *as a matter of fact*) or not (e.g., *whether this is a fact*). L1 and L2 speakers of English were presented with multi-word expressions and other phrases ending in the same word and Huang et al. used eye-tracking to measure fixation probabilities, first-fixation durations, and gaze durations. For their first experiment, they report that (the more predictable) words in multi-word expressions have significantly lower fixation probabilities and shorter first-fixation as well as gaze durations. A second, follow-up experiment on whether training would change the results for the L2 learner by making the final word of a multi-word expression more predictable and the results generally support that hypothesis as well.

The occasional utility of simple conditional probabilities notwithstanding, it has also been argued that a notable weakness of theirs is the absence of any degree of normalization or relativization. Consider Table 21.1 for a schematic 2×2 co-occurrence frequency table of the type that is widely used in corpus-linguistic studies (cognitive or otherwise). In this table, the element *E* of interest in the upper row might be a construction (e.g., the *as*-predicative) and the co-occurring element *X* might be a verb (e.g., *see*). Thus,
- the row total $a+b$ would be the frequency of the *as*-predicative in a corpus;
- the column total $a+c$ would be the frequency of *see* in a corpus;
- the cell a would be the frequency of *see* in the *as*-predicative.

TABLE 21.1 A schematic 2×2 co-occurrence frequency table

	Co-occurring element *X*	**Other elements (not *X*)**	**Totals**
Element *E*	a	b	$a+b$
Other elements (not *E*)	c	d	$c+d$
Totals	$a+c$	$b+d$	$a+b+c+d$

Quantifying the co-occurrence of *see* in the *as*-predicative on the basis of the conditional probability $^{a}/_{a+b}$ or $^{a}/_{a+c}$ neglects what happens in the other row (with $^{c}/_{c+d}$) or the other column ($^{b}/_{b+d}$). Based on that logic, corpus linguists have for decades preferred to express the association between *E* and *X* not just with conditional probabilities, but with association measures, which will be discussed in the following section.

2.1.2 Association Measures

The vast majority of association measures (AMs) in corpus-linguistic studies are based on tables of the kind exemplified in Table 21.1, which contain observed (co-)occurrence frequencies. While there has been a lot of debate on what is the right association measure, much of this debate is by now probably fairly fruitless because (i) it is likely that there simply is not one AM that fits all applications (more on that below) and (ii) the by far most frequently-used measures (i.e. the log-likelihood value G^2, (log) odds ratio (*OR*), pointwise *MI*, *t*, *z*, conditional probability $p(y|x)$, and ΔP) are actually all derivable from one and the same statistical approach, namely a simple logistic regression model that tries to predict, say, *X* from *E* (or vice versa).[2] The more important aspect that should be discussed is actually what the association measure reflects, a question that relates back to (i) above and that has, aside from some methodological articles, been examined too little: Does an AM only

- reflect association (like the odds ratio or ΔP) or does it also reflect the frequency of the element(s) in question? For instance, if one multiplied all frequencies in Table 21.1 by 10, does the AM change?
- consider one row/column of the table (the one containing cell *a*) or does its value also consider more information in the table (the other row/column or the column/row totals)?
- return a measure of mutual/bidirectional association between *E* and *X* or is the AM unidirectional and can, thus, distinguish the direction of association $E \rightarrow X$ from $X \rightarrow E$?

Considering the most widely-used AMs in terms of the above questions yields Table 21.2.

TABLE 21.2 A classification of the most widely-used association measures

	G^2	***OR***	p_{FYE}	***MI***	t	z	$p(y\|x)$	ΔP
AM reflects frequency and/or association?	f, a	a	f, a	a	f, a	f, a	a	a
AM considers not just row/column with *a*?	yes	yes	yes	yes	yes	yes	no	yes
AM is directional?	no	no	no	no	no	no	yes	yes

One of the maybe most widely used quantitative methods in quantitative corpus semantics is the family of methods referred to as collostructional analysis. Collostructional analysis as it has been used most of the time comprises three different methods of quantifying the co-occurrence preferences of words and/in constructions, all of which rely on some version of a 2×2 table such as Table 21.1:

- *collexeme analysis*, which quantifies the degree of attraction or repulsion of words (typically verbs) to a syntactically defined slot in a construction (see

2 An R script that shows how all these AMs are computed for a verb-construction frequency table of the kind used in collostructional analysis is available at ⟨http://www.stgries.info/research/2020_STG-PD_CooccData_PHCL.html⟩.

Stefanowitsch and Gries 2003), for example: how much does *see* like to occur in the *as*-predicative?
- (*multiple*) *distinctive collexeme analysis*, which quantifies which words (typically verbs) are attracted to or repelled by one of several constructions (see Gries and Stefanowitsch 2004a), for example: how much does *give* prefer to occur in the ditransitive as opposed to the prepositional dative?
- *covarying collexeme analysis*, which identifies preferred and dispreferred pairs in two slots of one construction (see Gries and Stefanowitsch 2004b), for example: the two verb slots in *The candidate tricked everyone into believing she was a linguist*.

In most applications, such a table is then statistically evaluated with the *p*-value of a Fisher-Yates exact test (p_{FYE} in Table 21.2 above)—an exact-test alternative to the approximate G^2 or a chi-squared tests—and discussed based on (i) the ratio of observed to expected *a* and (ii) the $\log_{10}$ of the *p*-value of FYE, which is interpreted as quantifying the degree to which *E* likes or dislikes to occur in/with *X*.

These methods have been applied in a variety of domains and languages including constructional senses and complementation patterns, syntactic alternations of a variety of constructions, verb-specific syntactic priming effects, analyses of diachronic changes in complementation patterns. One application of collexeme analysis is Gries, Hampe, and Schönefeld (2005), who study the *as*-predicative. They first perform a collexeme analysis on the construction, which returns *regard*, *describe*, *see*, *know*, and *treat* as the top 5 verb collexemes of the construction. They then validate the corpus results with a sentence-completion experiment in which subjects were presented with sentence fragments involving the 4 combinations resulting from crossing verbs that are frequent vs infrequent in the construction and verbs that are strongly vs. weakly attracted to the construction; another factor that the experimental design controlled for is the voice of the sentence fragment (given the construction's strong association to the passive). The results indicate that collostruction strength, not frequency, significantly predicts the frequency of subjects' *as*-predicative completions; Gries, Hampe, and Schönefeld (2010) provides similar converging evidence from a small self-paced reading study.

A more recent and interesting application is Perek (2014) which in fact involves an extension of collexeme analysis. His focus is on the verbs occurring in the conative construction (e.g., *John kicked at Mary*). Based on fictional prose data from the British National Corpus, he finds that even the most strongly attracted collexemes of this construction exhibit a considerable range of verbs/verb classes, which is at least unusual given that quite a few other collostructional studies of similar argument structure constructions have resulted in semantically much more homogeneous verb classes; the best example is probably the strong representation of transfer-related verbs in the ditransitive construction (see Stefanowitsch and Gries, 2003). Based on Croft's insightful critique of postulating constructional polysemy when all/most that motivates that notion is the occurrence of different verbs in a construction, Perek then does separate collexeme analyses on "sub-constructions" of the conative as defined by classes of verb senses (e.g., of cutting, pulling, or striking); once the resolution of the collexeme analysis is increased this way, the verbs preferred in the "sub-constructions" do indeed reflect their distinct notable semantic features.

Given its widespread application,[3] it seems fair to say that collostructional analysis is a useful way within cognitive-linguistic/usage-based semantics to implement the distributional hypothesis, i.e. the working assumptions of much corpus-linguistic work that has perhaps been formulated best by Harris (1970: 785f.):

> (i)f we consider words or morphemes A and B to be more different in meaning than A and C, then we will often find that the distributions of A and B are more different than the distributions of A and C. In other words, difference of meaning correlates with difference of distribution.

At the same time, collostructional analysis, as done so far, is inherently monofactorial: it studies the occurrence of some *X* given some *E*, that's it. Given the high degree of redundancy/overlap "built into" language, the approach yields good results, but it stands to reason that for many other applications, more factors or dimensions of information should be considered, which is why we now turn to such approaches.

2.2 *Multiple Variables 1: Multifactorial Predictive Modeling*

The first kind of multiple-variable approaches to consider involve a dependent/response variable whose conditional distribution given independent/predictor variables is explored (often to test hypotheses about which predictors are significantly correlated with the response). Since the values of the response variable are known—the data include the lexical/constructional/… choices speakers made—these are methods that fall under the heading of *supervised learning*.

3 This assessment is based on the fact that the two initial collostructions papers are both Stefanowitsch's and Gries's most-cited papers (at least according to Google Scholar, 8 March 2023).

2.2.1 Regression Modeling

One of the most frequent multifactorial approaches is regression modeling, with the vast majority of cases involving binary logistic regression modeling, i.e. scenarios where the dependent/response variable is binary (e.g., a choice between two words or two constructions) and the independent/predictor variables are numeric, ordinal, or categorical; see Hilpert and Blasi (2021) for an introductory article and Gries (2021: Chapters 5–6) for a textbook discussion.[4] One application of such modeling to a grammatical alternation—the realization vs. omission of *that* after *I think*—is a study by Shank, Plevoets, and Cuyckens (2014) of a stratified sample of approximately 5.8K instances of *think* with/without a complementizer in diachronic corpus data spanning the time period from 1560 to 2012. They annotated those instances for 26 predictors involving features of the corpus (file) as well as features regarding the matrix and the complement clause; the clause-based features involve, among others, person, tense, polarity as well as the length of material between the two clauses. They then perform a stepwise analysis to determine which predictors seem to affect *that* realization most. They find a variety of effects, in particular some interactions involving the predictor TimePeriod. For instance, over time *that* realization became less likely in spoken data, but more likely in written data. Similarly, the effect of the length of the complement subject or the harmony of polarity between matrix and complement clause are not constant/uniform across time.

Sokolova, Lyashevskaya, and Janda (2012) explore the locative alternation—the choice of a theme-object or a goal-object construction—of altogether eight prefixed and non-prefixed forms of the verb *gruzit'* ('load') based on approximately 1900 examples from the Russian National Corpus. Their predictors are the verb used (*gruzit'* vs. *nagruzit'* vs. *zagruzit'* vs. *pogruzit'*), whether the construction omits one participant (*no* vs. *yes*), and whether the verb is a participle. Their minimal adequate model has very high R^2 and C-scores (0.8 and 0.96) and indicates that (i) especially the verb lexeme is strongly predictive of the constructional choice (as one would expect if words and constructions interact within one constructicon) and that (ii) the prefixes of *gruzit'* are very unlikely to be semantically empty—minimally, their semantic contribution overlaps with that of the verb to which it is attached.

4 I am not sure I have ever seen an application in cognitive linguistics where researchers did not "downgrade" an ordinal predictor (e.g., a point on an animacy hierarchy or a complexity scale) to a categorical one, which is regrettable given the information loss it incurs; unfortunately, I myself have also done this when I shouldn't have.

Levshina, Geeraerts, and Speelman (2014), on the other hand, is an application of regression modeling to a lexical "alternation" / near synonymy kind of question, namely whether speakers would use *doen* or *laten*, i.e. causative constructions of the kind of *De politie deed/liet de auto stoppen*, "the police did/let the car stop," i.e. "the police stopped the car" and whether this lexical choice is co-determined by the directness of the causation involved. They annotated approximately 6.8K instances of both constructions for the semantic classes of the causer, causee, and caused event as well as the transitivity of the effected predicate and the affectedness of the causee. Then, they, too, analyzed the data with two logistic regression models involving *AIC*-based model selection, one with main effects only and one with interactions. The latter turns out to be superior and scores a good accuracy and, more importantly, *C*-score. They find that only some expectations of the (in)direct causation hypothesis are confirmed. They also proceed to discuss the issue of whether the results can be used to infer which instances of constructions are (close to instantiating) prototypes (following Gries, 2003a; 2003b) and how *doen* is quantitatively as well as semantically restricted, which they interpret as "*doen* seems to have more Gestalt-like semantics than *laten*, which has a looser set of semantic features" (p. 217).

Finally, based on the logic of lexically-specific effects of the kind uncovered in collostructional studies (see Section 2.1.2 above), they add an additional mixed-effects modeling analysis in which each effected predicate receives its own intercept adjustment (essentially following Baayen's 2011 suggestion). While the overall results are similar, they find that accounting for verb-specific effects this way (not unexpectedly) obviates the need for a topic-based predictor and they make a case for explaining such verb-specific effects as exemplar effects. Levshina et al. conclude *doen* is most likely in affective causation whereas the typical uses of *laten* can be captured best in a service frame.

While regression modeling is still probably the most widespread predictive-modeling technique, alternatives to it that are gaining in currency are a variety of tree-based approaches (such as classification and regression trees and random forests) as well as naïve discriminative learning, which I will turn to briefly now.

2.2.2 Alternatives to Regression Modeling

Over the last few years, machine-learning methods such as tree and forests have increased a bit in popularity in corpus linguistics in general, but now also in cognitive-linguistic approaches. One recent application is Fonteyn and Nini (2020), a study of whether gerunds are used with *of* (*eating of meat*) or not (*eating meat*). Approximately 14K instances from the EMMA (Early Modern

Multiloquent Authors) corpus are annotated for the response variable (realization vs. omission of *of*) and for three predictors (five kinds of determiner used and *none*, six functions of the gerund, and three verb types); in addition, they added the speaker producing the sentence as well as their age and generation and the genre of the text in which the gerund appeared. A conditional inference forest indicates that the language-internal predictor of determiner is by far the most important one (esp. with its levels *bare/no determiner* and *the*) and that that is true across nearly all individual speakers, but less important predictors vary a lot more between speakers.

One of the first presentations of naïve discriminative learning (NDL) in (cognitive) linguistics is Baayen (2011), largely a methodological paper comparing NDL against generalized linear mixed-effects models and classifiers (memory-based learning as well as support vector machines). One of the question Baayen starts out from (2011, 269f) is whether

> these different statistical models provide a correct characterization of the knowledge that a speaker has of how to choose between these two dative constructions. A statistical model may faithfully reflect a speaker's knowledge, but it is also conceivable that it underestimates or overestimates what native speakers of English actually have internalized.

The second, related question he is considering is how much and what kind of knowledge of frequency of (co-)occurrence speakers can be assumed to have. An NDL model is fit on the dative alternation and returns excellent *C*- and accuracy scores (0.97 and 0.92 respectively), a performance that is largely comparable to that of the other classifiers: in cross-validation, NDL performs slightly worse than mixed-effects models and support vector machines and slightly better than memory-based learning. Interestingly, NDL can be connected to the association measure ΔP (as used in collostructional studies, see above) and psychological theories of human learning (see Wagner and Rescorla, 1972) and achieves its good results without any researcher degrees-of-freedom.

2.3 *Multiple Variables 2: Multivariate* (*Exploratory*) *Approaches*

The second category of multiple-variable approaches also involves the consideration of many variables at the same time, but not necessarily with (the focus on) an obvious dependent/response variable—the focus is often more in identifying structure in the data. Since in many such cases the values of the response variable might not be known, several of these are methods that fall under the heading of *unsupervised learning*.

2.3.1 Behavioral Profiles

One approach towards especially (near) synonymy and polysemy that uses quantitative corpus data is that of behavioral profiles. This method involves the following steps:

- the retrieval of a sample of (ideally many) instances of the word(s) under consideration;
- the (usually) manual annotation of these instances for many features (called *ID tags*) from (usually) many levels of linguistic analysis, e.g. morphological, syntactic, semantic, lexical/collocational, and other features;
- the conversion of these data into vectors of percentages, the so-called *behavioral profiles*, such that for each word or sense under consideration, one obtains a percentage distribution of each ID tag of interest;
- the study of this table with descriptive and/or exploratory statistics (e.g., hierarchical cluster analysis).

In one of the first applications of this method, Gries (2006) studied the polysemy of the verb *run*. He annotates 815 examples of (all inflectional forms of) *run* (v.) from the British Component of the International Corpus of English and the Brown Corpus of American English for their senses (informed by dictionaries and WordNet) and 252 ID tags. He then demonstrates how the behavioral profile vectors can help address several usually thorny questions such as which sense is prototypical, whether to lump or split related senses, and where to connect senses in a (radial) network (of senses).

Gries's study is replicated by Glynn (2014b) on the basis of 500 occurrences of *run* from British and American English (conversation and online personal diaries). Glynn's findings largely corroborate Gries's earlier one, but Glynn also discusses potential follow-ups or improvements such as adding "social dimensions" (Glynn's cover term for the variables DIALECT and REGISTER) to the mix and also analyzing the data using correspondence analysis; he concludes with a plea towards meeting the challenges cognitive linguistics' conception of senses poses with methods that are capable of handling the resulting complexity of multivariate (corpus) data. (A final example of behavioral profiling will be discussed in the following section).

2.3.2 Cluster Analysis

Hierarchical cluster analysis is an exploratory, hypothesis-generating technique. It is normally used to group a set of elements (which could be examples, words, speakers, …) into clusters/groups such that the members of one group are very similar to each other and at the same time very dissimilar to members of other groups; see Moisl (2021) for a recent overview article and Moisl (2015) for a book-length introduction. In the case of hierarchical clustering, the user needs to specify (i) how the (dis)similarity of elements and the clusters/groups resulting from grouping together elements is quantified and (ii) how elements/clusters are to be merged. The result of such an analysis is typically a tree diagram revealing some structure in the clusterings of elements that can then, hopefully, be interpreted in an instructive/insightful way.

One application of hierarchical cluster analysis is Robinson (2014), a study of, among other things, how speakers differ in their uses of polysemous adjectives. She performs a hierarchical cluster analysis on the altogether 35 meanings of 8 adjectives based on the frequencies of use by 72 speakers. Assuming (in this paper) a 3-cluster solution and validating her results with inferential statistics (logistic regression and classification trees), she finds that the clusters of sense frequencies of the adjectives are indeed strongly and predictively correlated with generational differences and socio-demographic factors.

Another interesting cluster-analytic case study is Desagulier (2014), who explores the use of degree modifiers (such as *rather*, *quite*, *fairly*, and *pretty*) as a function of the adjectives they modify. While he starts out from a simple but massive concordance of these and 19 other modifiers in the Corpus of Contemporary American English—the concordance returned more than 316k co-occurrence tokens involving 432 different modified adjectives—the cluster analysis he uses is not applied to "mere" co-occurrence frequencies, but to the collexeme strengths as determined by a set of per-modifier collexeme analyses (see Gries and Stefanowitsch, 2010 for the first such application). Following Divjak and Gries (2006), he computes a 23 × 23 dissimilarity matrix using the Canberra metric for the similarities of the modifiers based on collexeme strengths; then he uses Ward's method to amalgamate the modifiers into a cluster tree/dendrogram; finally, he computes bootstrapping-based cluster significance values. As a result, he obtains four very well "functionally and semantically motivated" (p. 164) clusters: one with maximizers, one with diminishers, one with moderators, and one with boosters, a result he considers as partial support of earlier work on the synonymy of moderators.

Yet another "more involved" application of cluster analysis is on a set of behavioral profile vectors as in the pair of papers of Divjak and Gries (2006; 2008). In the former, they report on the results of a behavioral profile analysis of approximately 1.6k sentence featuring nine Russian verbs meaning 'to try.' The 1.6k instances were annotated for altogether 87 morphological, syntactic, and semantic ID tags and submitted to a hierarchical cluster analysis, which returned three groups of near synonyms. These were then analyzed with regard to the differences between clusters as well as the differences between verbs within one and the same cluster (using pairwise differences of ID tag percentages and *t*-/*F*-scores). The between-cluster differences can be summarized as follows (from Divjak and Gries 2008:193f.):

- a human is exhorted to undertake an attempt to move himself or others (rather than to undertake mental activities); often, these activities are negated;
- an inanimate subject undertakes repeated non-intense attempts to exercise physical motion; the actions are often uncontrollable and fail;
- an inanimate subject (concrete or abstract) attempts very intensely but in vain to perform what typically is a metaphorical extension of a physical action.

In order to validate the corpus-based findings, Divjak and Gries (2008) analyze the outcome of a series of sorting experiments with native speakers of Russian who were asked to sort nine sentences that only differed in their verb meaning 'to try' into groups base on their overall semantic similarity. Then, they computed a score quantifying the fit between the cluster analysis of the observational/corpus data and the cluster analysis of the experimental sorting data, which was then compared to the range of scores one might obtain from a null hypothesis distribution. The results show that the speakers' sorting solutions are very consistent with the corpus-based cluster-analytic results; similarly supportive results were obtained from a comparison of the corpus-based clustering to an identical cluster analysis of the experimental data and a gap-filling task. This study is methodologically interesting in how cluster analyses from observational and experimental data are compared and evaluated.

2.3.3 Correspondence Analysis

Correspondence analysis is an exploratory statistical method in spirit not at all unlike principal component/factor analysis (or multidimensional scaling) to discover patterns in two- or higher-dimensional frequency tables based on how row and column frequencies and residuals pattern with regard to each other; as in other dimension reduction techniques, the result of a correspondence analysis is a 2- or 3-dimensional plot (that is actually very easy to *mis*interpret, which might be one reason for the relative rarity of this method); see Glynn (2014c) for an overview.

One application of a correspondence analysis is Delorge, Plevoets, and Colleman (2014). They study the corpus frequencies with which dispossession verbs with *ont-* 'away' occur in a variety of possessional transfer constructions. In a synchronic analysis, they find that verbs fall into a number of clusters based on the constructions they (do not) 'like' to occur in that exhibit clear patterns in terms of their semantics and in terms of which participants are lexically profiled/realized. In a diachronic analysis, on the other hand, they find evidence for constructional specialization such that, over time, constructions solidify their preferences for certain constructions.

Another application in the same volume is the already mentioned Desagulier (2014). After his initial cluster analysis of the separate collexeme analyses of 23 degree modifiers, he also computes a correspondence analysis of degree modifiers based on their highest co-occurrence frequencies with the adjectives they modify. Some of the results are amazingly clear-cut especially once the collexemes are grouped into semantic classes: For instance, the collexemes of *pretty* and *quite* form very clearly delineated clouds, and *pretty* and *quite* combined are also quite different from *fairly* and *rather* combined, where the latter two are particularly well distinguished by *rather*'s negative semantic prosody; the following list is quoted from his paper (p. 1762):

- *rather*: dimension or position in space (e.g., *long*, *high*), atypicality/oddity (e.g., *odd*, *bizarre*), negative attitudes (e.g., *ironic*), unclearness (e.g., *vague*, *obscure*);
- *quite*: epistemic, dynamic, and factual meanings (e.g., *likely*, *able*, *true*), difference (e.g., *different*, *separate*), psychological states (e.g., *surprised*, *concerned*, *content*);
- *fairly*: location in time (e.g., *recent*, *new*), typicality (e.g., *typical*, *common*, *standard*);
- *pretty*: appreciative and unappreciative values (e.g., *good*, *great* vs. *bad*, *awful*), cleverness and stupidity (e.g., *smart* vs. *stupid*, *dumb*), difficulty (e.g., *difficult*, *tough*, *hard*), psychological stimuli (e.g., *scary*, *funny*).

A final and very interesting example is Flach (2020), who studies *go*/*come* (*and*) V constructions in various syntactic environments (e.g. imperative, indicative, etc.) based on data from the Corpus of Contemporary American English. A correspondence analysis identifies that two dimensions account for 88.6% of the structure in the data and, more particularly, even returns an assertive-directive continuum (from *do* to the imperative). She then also conducts an acceptability judgment experiment of stimuli whose verbs after *go*/*come* were determined on the basis of collexeme analyses. The judgments (z-standardized within each participant) resulting from the experiment were then analyzed with a linear mixed-effects model to determine to what degree they are correlated with the syntactic environment and the collexeme association with the construction. Intriguingly, she finds that the acceptability ratings are strongly correlated with the results from the correspondence analysis whereas there is no such strong relation with the frequencies in the construction; also, the corpus-based results are robust (as determined by comparisons with other corpora/registers).

3 New Developments and Desiderata

As the previous sections have hopefully illustrated, the field has evolved considerably from a relative absence of quantitative studies to a plethora of different quantitative multifactorial and multivariate methods that often combine observational and experimental data in insightful way. But it doesn't end there because some studies have gone even beyond that kind of versatility and have enriched the many "traditional" modeling techniques already in use in cognitive linguistics/semantics with methods from other fields and I want to discuss two such approaches I consider particularly interesting.

3.1 *Network Approaches*

The first such approach involves the use of (social) network analysis and one particular comprehensive study—admittedly a book-length treatment—is Ellis et al (2016). Their monograph is a very detailed study of the acquisition, use, and transmission of verb-argument constructions (VACs) based on corpus and experimental data, but on top of all that they also explore the semantic associations between verbs and VACs using semantic graphs/networks (built from WordNet's synsets for verbs). This kind of network analysis is particularly interesting in how (i) these networks can be built with relatively little researcher input with regard to semantics (i.e., fewer researcher degrees of freedom) and how (ii) they yield results that inform many central semantic questions including prototypical members of (semantic) categories, the coherence of categories, polysemy detection, and others, all based on methods/statistics from network analysis involving degree/betweenness centrality, community detection etc. For example, their analysis of the V-*about*-N VAC returns eight clusters/communities that reflect clearly coherent senses such as communication expression, communication reception, cognition concern, physical movement in space, to name a few examples. This kind of work is promising both for its empirical rigor and its integratability with more traditional methods (as exemplified throughout all of Ellis et al., 2016).

3.2 *Inductive and Deep Learning Approaches*

Another class of approaches that is currently emerging in cognitive semantics are ones that might informally well be called "high-powered computational (learning) methods," methods that are less "traditional statistical methods or models" but more computational inductive and/or deep learning methods.

3.2.1 An Extension of Association Measures

One approach that extends the logic of association measures from Section 2.1.2 above involves the automatic learning/identification of constructions as in, for example, Dunn (2017). Dunn's construction induction algorithm is based on a combination of "linguistic resources" (e.g., a part-of-speech tagger, a semantic analysis system, and a dependency parser) and "mathematical modeling resources" involving frequency counts and association measures (specifically the directional association measure ΔP extended to work with multi-unit units); as per Dunn (2017:266), which is worth quoting at length:

> The construction induction algorithm is based on multi-directional (left-to-right or right-to-left), multi-dimensional (across varying levels of representation), multi-length (across two or more units) association strength, measured with and without complex constituent-internal structure (i.e., distance is measured at different levels of abstraction). The idea is that sequences which are constructions (e.g., are cognitively entrenched to some degree) are more internally associated than sequences which are not constructions (e.g., those which are chance co-occurrences of units). The purpose of the association measures (and the frequency counts on which such measures are ultimately based) is to learn an inventory of constructions from the very large hypothesis space of all observed sequences.

For evaluation, the proposed algorithm is run on 1 billion words/40 million sentences from the ukWac web-based corpus and Dunn discusses several constructions returned by it including those in (1) and (2).

(1) a. *wh*-determiner + modal + *be* + past participle
b. that will be provided
c. that should be made

(2) a. *to* + verb + determiner + noun
b. to get an idea
c. to sell a product

Dunn also discusses limitations of this approach, i.e. constructions with "incorrect boundaries" and the more limited degree to which his algorithm reflects psycholinguistic reality on the level of a speaker. Particularly promising characteristics of the algorithm are its stability both with regard to the consistency of (i) the coverage of constructions and (ii) the stability across differently-sized data sets and makes a compelling case for one of the central working assumptions of cognitive linguistics, that a grammar can be learned from the input even in absence of a universal innate grammar module; see Beekhuizen and Bod (2014) for an earlier interesting exploration of unsupervised construction identification.

3.2.2 Distributional Semantics

One of the most recent developments in cognitive semantics involves the use of vector-space semantics and deep learning algorithms such as classical vector-space semantics of the type discussed in Manning and Schuetze (1999: Chs. 8, 15) or Jurafsky and Martin (2020: Ch. 6), but also newer techniques—deep learning models trained on vast amount of texts from which they acquire co-occurrence information of many linguistic kinds—such as word2vec (Mikolov 2013; Mikolov et al., 2013), GloVe (Pennington et al., 2014), or BERT (Devlin et al., 2018).

An example of a more traditional vector-space semantic analysis is Perek and Hilpert's (2017) tweaking of Gries and Hilpert's Variability-based Neighbor Clustering (Gries and Hilpert, 2008) to work with vector-space representations to study the diachronic development of constructions (such as V *the hell out of* NP construction and the V POSS *way* PP construction). For the former, fairly new construction, their 1930s to 2000s data from the Corpus of Historical American English reveal a slow and gradual expansion; for the latter, the data are noisier but are interpreted as a three-time-periods solution, with each period featuring somewhat distinctive verbs in the *way* construction; see Perek (2018) for an interesting follow-up to this study and Kutuzov et al. (2018) for a recent survey.

Another example of an interesting vector-space application is Levshina and Heylen's (2014) work on Dutch causative constructions; an at least somewhat related approach is Gries (2018), who defines constructional prototypes for constructional alternations and then showcases the high degree of predictive power of deviations from those prototypes (measured using the Kullback-Leibler divergence).

Studies involving the newer algorithms—BERT, fastText, etc.—are still rare in cognitive linguistics, which is not surprising, given their recency. One very recent application is Madabushi et al. (2020), who explore to what degree word embedding models such as BERT acquire constructional knowledge from texts and, therefore, are able to identify constructions; they conclude that initial results are promising: there is a tendency for BERT to return sentences as constructions that construction grammarians would consider constructions (but also many patterns that construction grammarians would probably not consider constructions).

4 Concluding Remarks

While the increased use of quantitative methods in cognitive linguistics has been welcomed by many (see, e.g., Glynn 2014a; Janda, 2017: 511–512), there have also been naysayers; see Divjak, Levshina, and Klavan (2016: 453) for mentions of "concerns" that have been raised. However, I think those concerns are exaggerated and in part biased, given that only ever hears concerns about too much empiricism but never concerns about too much theory. To my mind at least, a theory's plausibility is a function of how well it can account for empirical data or make predictions on to-be-collected empirical data—a theory/theoretical model that does not come with the implied commitment to make testable predictions probably does not do much to advance a field. And since a statistical "model is a formal representation of a theory" (Adèr 2008: 280), it is with statistical modeling (or tools) that we test theories (at least if they are sufficiently precise in their predictions, which is of course a different question). This is especially relevant when cognitive linguists deal with phenomena where a linguistic choice is co-determined simultaneously by literally dozens of contextual, phonological, lexical, semantic, structural, information-structural, psycholinguistic, and sociolinguistic predictors—if statistics were taboo, intuition would not be enough for that. As Glynn (2014a: 16–18) shows (and see of course the groundbreaking classic Sandra and Rice 1995), cognitive linguists failed to agree even on the numbers of senses of even the most overstudied lexical elements, and he argues, correctly, I believe, that it is impossible to understand how all formal and functional/semantic dimensions interact. I also agree with Glynn (2014a: 7), who says "given the theoretical assumptions of Cognitive Linguistics, it is argued that quantitative corpus-driven methods are essential for the description of semantic structures" and therefore hope that quantitative methods are here to stay;[5] as Arppe et al. (2010:3) state, "The benefits of multi-

5 See Jenset and McGillivray (2017: Section 3.7) for a wonderful series of arguments against quantitative naysayers; they make this point much better than I could ever have. It should go without saying that endorsements of quantitative methods come with of course all the usual caveats: they need to be done right (both in terms of how the chosen method fits the study's goal and in terms of the requirements of the method *per se*) and they need to be reported on at a level of resolution that ensures replicability.

methodological research outweigh the problems—in linguistics as much as elsewhere."

References

Adèr, Hermanus. 2008. Modelling. In H.J. Adèr and G.J. Mellenbergh (eds.), *Advising on Research Methods: A Consultant's Companion*, 271–304. Huizen: Johannes van Kessel.

Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert and Arne Zeschel. 2010. Cognitive Corpus Linguistics: five points of debate on current theory and methodology. *Corpora* 5(1): 1–27.

Aslin, Richard N., Jenny R. Saffran and Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9(4): 321–324.

Baayen, R. Harald. 2011. Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics* 11(2): 295–328.

Beekhuizen, Barend and Rens Bod. 2014. Automating construction work: Data-Oriented Parsing and constructivist accounts of language acquisition. In R. Boogaart, T. Colleman, and G. Rutten (eds.), *Extending the scope of Construction Grammar*, 47–74. Berlin: De Gruyter.

Bybee, Joan and Paul J. Hopper (eds.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Bybee, Joan and Sandra A. Thompson. 1997. Three frequency effects in syntax. *Berkeley Linguistics Society* 23: 65–85.

Delorge, Martine, Koen Plevoets, and Timothy Colleman. 2014. Competing "transfer" constructions in Dutch: the case of *ont*-verbs. In D. Glynn and J.A. Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 39–60. Amsterdam: John Benjamins.

Desagulier, Guillaume. 2014. Visualizing distances in a set of near synonyms: rather, quite, fairly, and pretty, In D. Glynn and J.A. Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 145–178. Amsterdam: John Benjamins.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Available at arXiv:1810.04805v2.

Divjak, Dagmar S. and Stefan Th. Gries. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1): 23–60.

Divjak, Dagmar S. and Stefan Th. Gries. 2008. Clusters in the mind? Converging evidence from near synonymy in Russian. *The Mental Lexicon* 3(2): 188–213.

Dunn, Jonathan. 2017. Computational learning of construction grammars. *Language and Cognition* 9: 254–292.

Ellis, Nick C., Ute Römer, and Matthew Brook O'Donnell. 2016. Usage-based approaches to language acquisition and processing: cognitive and corpus investigations of construction grammar. *Language Learning* 66(S1): 1–358.

Flach, Susanne. 2020. Schemas and the frequency/acceptability mismatch: Corpus distribution predicts sentence judgments. *Cognitive Linguistics* 31(4): 609–645.

Fonteyn, Lauren and Andrea Nini. 2020. Individuality in syntactic variation: an investigation of the seventeenth-century gerund alternation. *Cognitive Linguistics* 31(2): 279–308.

Glynn, Dylan. 2014a Polysemy and synonymy: cognitive theory and corpus method. In D. Glynn and J.A. Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 7–38. Amsterdam: John Benjamins.

Glynn, Dylan. 2014b. The many uses of *run*: Corus methods and socio-cognitive semantics. In D. Glynn and J.A. Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 117–144. Amsterdam: John Benjamins.

Glynn, Dylan. 2014c. Correspondence analysis: exploring data and identifying patterns. In D. Glynn and J.A. Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 443–485. Amsterdam and Philadelphia: John Benjamins.

Gries, Stefan Th. 2003a. *Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement*. London: Continuum Press.

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

Gries, Stefan Th. 2006. Corpus-based methods and cognitive semantics: the many meanings of *to run*. In S.Th. Gries and A. Stefanowitsch (eds.), *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*, 57–99. Berlin: Mouton de Gruyter.

Gries, Stefan Th. 2018. The discriminatory power of lexical context for alternations: an information-theoretic exploration. *Journal of Research Design and Statistics in Linguistics and Communication Science* 5(1–2): 78–106.

Gries, Stefan Th. 2019. *Ten Lectures on Corpus-linguistic Approaches: Applications for Usage-based and Psycholinguistic Research*. Leiden: Brill.

Gries, Stefan Th. 2021. *Statistics for linguistics with R: a practical introduction* (3rd rev. and ext. ed.). Berlin: De Gruyter Mouton.

Gries, Stefan Th. and Dagmar S. Divjak. 2010. Quantitative approaches in usage-based cognitive semantics: myths, erroneous assumptions, and a proposal. In D. Glynn and K. Fischer (eds.), *Quantitative Methods in Cognitive Semantics: Corpus-driven Approaches*, 333–354. Berlin: Mouton de Gruyter.

Gries, Stefan Th., Beate Hampe and Doris Schönefeld. 2005. Converging evidence:

Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4): 635–676.

Gries, Stefan Th., Beate Hampe and Doris Schönefeld. 2010. Converging evidence, volume II: More on the association of verbs and constructions. In S. Rice and J. Newman (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*, 59–72. Stanford: CSLI.

Gries, Stefan Th. and Martin Hilpert. 2008. The identification of stages in diachronic data: variability-based neighbor clustering. *Corpora* 3(1): 59–81.

Gries, Stefan Th. and Anatol Stefanowitsch. 2004a. Extending collostructional analysis: a corpus-based perspective on "alternations." *International Journal of Corpus Linguistics* 9(1): 97–129.

Gries, Stefan Th. and Anatol Stefanowitsch. 2004b. Co-varying collexemes in the into-causative. In M. Achard and S. Kemmer (eds.), *Language, Culture, and Mind*, 225–236. Stanford: CSLI.

Gries, Stefan Th. and Anatol Stefanowitsch. 2010. Cluster analysis and the identification of collexeme classes. In S. Rice and J. Newman (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*, 73–90. Stanford: CSLI.

Harris, Zellig S. 1970. *Papers in Structural and Transformational Linguistics*. Dordrecht: Reidel.

van Heuven, Walter J.B., Pawel Mandera, Emmanuel Keuleers and Marc Brysbaert. 2014. SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology* 67(6): 1176–1190.

Hilpert, Martin and Damián E. Blasi. 2021. Fixed-effects regression modeling. In M. Paquot and S.Th. Gries (eds.), *A Practical Handbook of Corpus Linguistics*. Berlin: Springer.

Huang, Ping-Yu, David Wible and Hwa-Wei Ko. 2012. Frequency effects and transitional probabilities in L1 and L2 speakers' processing of multiword expressions. In S.Th. Gries and D.S. Divjak (eds.), *Frequency Effects in Language Learning andProcessing*, 145–175. Berlin: De Gruyter Mouton.

Janda, Laura A. 2017. The quantitative turn. In Barbara Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*, 498–514. Cambridge: Cambridge University Press.

Jenset, Gard B. and Barbara McGillivray. 2017. *Quantitative historical linguistics*. Oxford: Oxford University Press.

Jurfasky, Daniel and James H. Martin. 2020. *Speech and Language Processing* (3rd ed.). Draft of 03 Dec 2020 Available at URL ⟨https://web.stanford.edu/~jurafsky/slp3/⟩, accessed 12 March 2021.

Kutuzov, Andrey, Lilja Øvrelid, Terrence Szymanski and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: a survey. *Proceedings of the 27th International Conference on Computational Linguistics* (*COLING2018*): 1384–1397.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford: Stanford University Press.

Levshina, Natalia, Dirk Geeraerts and Dirk Speelman. 2014. Dutch causative constructions: Quantification of meaning and meaning of quantification. In D. Glynn and J.A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 205–221. Amsterdam: John Benjamins.

Levshina, Natalia and Kris Heylen. 2014. A radically data-driven Construction Grammar: Experiments with Dutch causative constructions. In R. Boogaart, T. Colleman and G. Rutten (eds.), *Extending the Scope of Construction Grammar*, 17–46. Berlin: De Gruyter Mouton.

Levshina, Natalia. 2021. Conditional inference trees and random forests. In M. Paquot and S.Th. Gries (eds.), *A Practical Handbook of Corpus Linguistics*. Berlin: Springer.

Madabushi, Harish Tayyar, Laurence Romain, Dagmar S. Divjak and Petar Milin. 2020. CxGBERT: BERT meets Construction Grammar. COLING 2020.

Manning, Christopher D. and Hinrich Schuetze. 1999. *Foundations of statistical natural languge processing*. Cambridge, MA: The MIT Press.

Mikolov, Tomas. 2013. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*. Available at DOI: arXiv:1310.4546.

Mikolov, Tomas, Kai Chen, Greg Corrado and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. Available at DOI: arXiv:1301.3781.

Moisl, Hermann. 2015. *Cluster analysis for corpus linguistics*. Berlin: De Gruyter Mouton.

Moisl, Hermann. 2021. Cluster analysis. In M. Paquot and S.Th. Gries (eds.), *A Practical Handbook of Corpus Linguistics*. Berlin: Springer.

Pennington, Jeffrey, Richard Socher and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), 1532–1543.

Perek, Florent. 2014. Rethinking constructional polysemy: the case of the English conative construction. In D. Glynn and J.A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 61–85. Amsterdam: John Benjamins.

Perek, Florent. 2018. Recent change in the productivity and schematicity of the *way*-construction: a distributional semantic analysis. *Corpus Linguistic and Linguistic Theory* 14(1): 65–97.

Perek, Florent and Martin Hilpert. 2017. A distributional semantic approach to the periodization of change in the productivity of constructions. *International Journal of Corpus Linguistics* 22(4): 490–520.

Rescorla, Robert A. and Allan R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black and

W.F. Prokasy (eds.), *Classical conditioning*, volume II: *Current theory and research*, 64–99. Appleton-Century-Crofts.

Robinson, Justyna A. 2014. Quantifying polysemy in cognitive sociolinguistics. In D. Glynn and J.A. Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 87–115. Amsterdam: John Benjamins.

Sandra, Dominiek and Sally Rice. 1995 Network analyses of prepositional meaning: mirroring whose mind—the linguist's or the language user's? *Cognitive Linguistics* 6(1): 89–130.

Shank, Christopher, Koen Plevoets and Hubert Cuyckens. 2014. A diachronic corpus-based multivariate analysis of "I think that" vs. "I think zero." In D. Glynn and J.A. Robinson (eds.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, 279–303. Amsterdam: John Benjamins.

Sinha, Chris. 2017. Opening commentary: getting the measure of meaning. In B. Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*, 493–497. Cambridge: Cambridge University Press.

Sokolova, Svetlana, Olga Lyashevskaya and Laura A. Janda. 2012. The Locative Alternation and the Russian "empty" prefixes: A case study of the verb gruzit' "load." In D.S. Divjak and S.Th. Gries (eds.), *Frequency Effects in Language Representation*, 51–86. Berlin: De Gruyter Mouton.

Stefanowitsch, Anatol and Stefan Th. Gries. 2003. Collostructions: investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2): 209–243.