Dylan Glynn*

Quantifying polysemy: Corpus methodology for prototype theory

DOI 10.1515/flin-2016-0016
Submitted August 31, 2015; Revision invited February 20, 2016;
Revision received March 10, 2016; Accepted May 31, 2016

Abstract: This study addresses the methodological problem of result falsification in Cognitive Semantics, specifically in the descriptive analysis of semasiological variation, or "polysemy." It argues that manually analysed corpus data can be used to describe models of semantic structure. The method proposed is quantified, permitting repeat analysis. The operationalisation of a semasiological structure employed in the study takes the principle of semantic features and applies them to a contextual analysis of usage-events, associated with the lexeme under scrutiny. The feature analysis, repeated on a large collection of occurrences, results in a set of metadata that constitutes the usage-profile of the lexeme. Multivariate statistics are then employed to identify patterns in those metadata. The case study examines 500 occurrences of the English lexeme *annoy*. Three basic senses are identified as well as a more complex array of semantic variations linked to morpho-syntactic context of usage.

Keywords: *annoy*, corpus methods, multivariate statistics, polysemy, prototype theory, Usage-Based Model

1 Introduction: The Usage-Based Model and Cognitive Semantics

1.1 Problem: Result falsification and the Usage-Based Model

Arguably, one of the most important qualities of a scientific description is that the method employed allows for the possibility of falsifying the results it produces. Typically, this takes the form of a repeat analysis of a new sample. Whether one is testing a theory about language functioning or performing a purely descriptive study of a specific language, the need to be able to falsify

***Corresponding author: Dylan Glynn,** DEPA – University of Paris VIII, 2 Rue de la Liberté, 93526 Saint-Denis, France, E-mail: dglynn@univ-paris8.fr

results is essential. This study addresses the question of how this can be achieved for a semantic analysis which adopts the Usage-Based Model of language.

Within the development of the theoretical paradigm of Cognitive Linguistics (Fillmore 1985; Talmy 1985; Lakoff 1987; Langacker 1987), semasiological variation and the phenomenon of polysemy have played a central role. Originally, Lakoff (1973, 1987), Fillmore (1975, 1985), Coleman and Kay (1981), Geeraerts (1986), and Taylor (1989) argued that the truth-conditional approaches to semantic structure lead to inadequate descriptions of language. This premise is taken as axiomatic in the paradigm. However, to find fault with an existing theory is one thing; it is another to offer an alternative. Drawing on fuzzy set theory in mathematics (Zadeh 1965, 1968; Goguen 1967, 1969) and prototype set theory in psychology (Heider 1971, 1972; Rosch 1973, 1975), the above-mentioned authors argued that the solution to describing and explaining semantic structure lies in the modelling of categories that are structured by non-discrete boundaries and graded membership.

Despite the productivity of what we may term Radial Set Analysis in the polysemic modelling that ensued, a fundamental weakness was identified.[1] Sandra and Rice (1995) performed an elicitation-based study demonstrating that Radial Set analyses offered no way to empirically confirm findings. First, in Radial Set Analysis, the categories analysed were often subjective, and the data were entirely introspective. Although arguably not problematic in itself, these two points can lead to issues in determining the descriptive adequacy of a given study. Second, the categories were often based on relative membership, so that negative evidence (in the form of counterexamples) could not be used to falsify the results.[2] This second issue is a fundamental limitation of all inductive research, where significant patterns are proposed instead of discrete rules or "laws." Indeed, the issue reveals what is arguably the biggest problem for Radial Set Analysis: The principle failing of the research tradition was the method of analysis rather than the theoretical model it was designed to confirm.

Formal semantics, and the tests it employs, relies on deductive evidence. In principle, this means that a single possible counter example can disprove a

[1] The term Radial Set Analysis was coined by Geeraerts (1989: 9), but the approach is known under various names. It was especially influential and productive in the 1980s and 1990s. Exemplary studies include Brugman (1983), republished as the study of *over* in Lakoff (1987), Rudzka-Ostyn (1985), Vandeloise (1986), Herskovits (1986), Janda (1990), and Cuyckens (1995). See Glynn (2014a) for a summary of this research tradition.

[2] See Wierzbicka (1990) for more theoretical discussion of problems along these lines.

hypothesis. In contrast, cognitive and functional approaches assume the Usage-Based Model of language (Hopper 1987; Langacker 1987) rendering this kind of deductive evidence irrelevant to hypothesis testing. The model holds that language structure (*langue*/competence) is a result of usage (*parole*/performance), not the reverse (as propounded by the Structuralist – Generative approaches to which formal semantics adheres). Put simply, from the perspective of the Usage-Based Model, structure is an epiphenomenal effect, a generalisation across usage events, yet the mainstream of Radial Set Analysis continued to employ methods best suited to the competence-based/rule-based Structuralist model of language.

Given a usage-based understanding of language, individual events/instantiations cannot prove or disprove a hypothesis; the truth of a hypothesis can only be demonstrated inductively. In other words, just like most of empirical science, usage-based linguistics needs to work with samples, extrapolate from those samples, and compare predictions with observations. Technically, inductive research does not involve proofs but predictive accuracy. First, a hypothesis should be able to account for all observations. This is self-evident and *sine qua non*. Second, observations that support a hypothesis must be shown to be "significant," understood as "probably" representative of the entire population (for linguists, that is language). Probable is determined by a quantified degree of confidence – in social science, this is typically a 5% probability that the hypothesis is false and in natural science a 1% probability. Seen from this perspective, the case studies in the Radial Set approach to polysemy were in fact hypotheses, theoretical models of the semasiological structure of lexemes and morphemes, rather than actual case studies *per se*. The method of analysis used, a combination of semantic criteria and introspective analysis, did not conform to the methodological requirements of the theory it sought to confirm. A primary aim in this study is to show how empirical observational data can be used to describe polysemy, assuming a theory of graded non-discrete set membership and the Usage-Based Model.

1.2 Problem: *A-priori* senses and the Usage-Based Model

The notions of prototype-structured sets and fuzzy bounded sets were originally applied to extra-lexical denotata or to the onomasiological choices in categorising denotata into predetermined designata. Although also applicable to semasiological categorisation, in the era of Radial Set Analysis, the full implications of adapting these notions for the empirical description of polysemy were, arguably, not fully appreciated. With no formal/discrete anchor (such as a lexeme or a morpheme), as is found in onomasiological research, it is far from clear what actually constitutes the categories under investigation. In other words, how can we determine what the category of lexical sense is so that this category can be subject to (or, indeed, itself determine) prototype membership or fuzzy boundedness? Radial Set Analysis often treated senses as "nodes" in a network; a metaphor that reifies, in a discrete manner, the notion of "sense." Yet within Cognitive Linguistics, even at the time, some linguists were questioning the appropriateness of thinking about meaning in such terms. Consider a quote from Geeraerts (1993a) on the nature of polysemy, from the era of Radial Sets:

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

Although, if asked, cognitive linguists would likely agree that lexical senses are not discrete categories and that they cannot be reified except for purposes of discussion, Radial Set Analysis offers no ready means for describing lexical structure except in terms of reified discrete categories. Again, it appears that method was not adapted to theory.

In Radial Set Analysis, Lakoff (1987: 420) made explicit the idea of schematic features, where configurations of various features of an abstract schema are understood as a lexical sense. As a theoretical model of the semasiological structure of a lexeme, this is unproblematic. The question remains as to how to extend the principle to individual instances, or tokens. Since Lakoff (1987) and others were working with a unique invented example per configuration, the need to see these sets of features (determining a single sense) as relative in themselves did not arise. In a usage-based study, applying prototype and fuzzy set theory to semasiological structure involves evoking the graded set theory twice, at two levels of analysis: (i) the relationship between the different senses with respect to the lexeme; (ii) the relationship between the different uses with respect to the lexical sense (to the extent that one believes a sense exists). Radial Set Analysis focuses on the use of graded set membership to talk about the relationship between individual senses, but empirically, actual instances of use are also concerned with the categorisation of individual posited senses. This kind of issue was addressed in depth by Geeraerts (1986, 1990), who summarises the problem succinctly:

> [O]n the one hand, there is the analytical attempt to define subsets of the observed [instantiations] of [the lexeme], on the other, there are intuitive observations as to the distinctions and similarities between those applications. (Geeraerts 1990: 204)

In order to apply Radial Set Analysis to observable data, introspection-based semantic categories (senses/nodes/types) themselves must be first posited and then, again using introspection-based semantic criteria, the observed instances (examples/occurrences/tokens) must be categorised.

Geeraerts (1986, 1990, 1995) and Rudzka-Ostyn (1988, 1989, 1995) worked with various heuristics and representational formats to overcome the problem that results from working with samples of natural data, where any given occurrence may or may not possess all the features of a given configuration (lexical sense/radial node/observable type). The descriptive tools they developed work well in near-synonymy studies, such as Dirven et al. (1982), Rudzka-Ostyn (1988, 1995), Geeraerts et al. (1994). In such onomasiological research, the categories being investigated exist *a priori* and are necessarily discrete, thus eliminating the need to posit hypothetical semasiological categories/lexical senses. Generalising the principle of feature analysis to semasiological structuring, without the notion of discrete lexical senses, is difficult.

A first question that presents itself is determining the features and adequately operationalising them. The more abstract the concept at hand, the more difficult this becomes. In the case of spatial prepositions, where a Landmark, a Trajector, and possibly a Path are literally or figuratively present, the selection of features is relatively straightforward. This, however, is not the case for all lexical concepts. Second, although a lexical sense may be operationalised as a configuration of semantic features, as was proposed by Lakoff (1987) for the case of *over*, does one begin in such an analysis with features believed to elucidate the semasiological structure or does one begin with the senses themselves and then look for features that will distinguish them. This is a non-trivial question and will likely result in divergent results, one favouring a maximally fine-grained network and the other a set of more coarse-grained distinctions. Third, given the Usage-Based Model of language, it makes the most sense to begin with the data and build up to lexical senses, based upon the “observed” clustering of features rather than with a “configuration” of features posited *a priori*. Such a bottom-up, or *a-posteriori*, approach would solve the above-mentioned problems, but how to determine the features and identify such clusters are difficult questions in themselves.

This study takes this bottom-up approach. In order to answer the question of identifying semantic features for abstract concepts, it draws particularly on Rudzka-Ostyn’s (1989) work. Her research in polysemy specifically developed the usage-event-based features, such as actor types and their relations. In order to identify the clusters believed to index senses, the study draws on the quantitative tradition developed in Multifactorial Usage-Feature Analysis, described below.

1.3 Proposal: Usage-Feature Analysis, behavioural profiles, *a-posteriori* senses

Within Cognitive Linguistics, especially in research on near-synonymy, an analytical method has been developed that can arguably resolve some of the problems identified above. A series of studies, Dirven et al. (1982), Rudzka-Ostyn (1988, 1995), Geeraerts et al. (1994, 1999), and Gries (1999, 2003), demonstrated that the traditional semantic analysis of usage-features of a given linguistic phenomenon, if performed across large numbers of naturally occurring instances of that phenomenon, produces a usage-profile. This usage-profile, if one acknowledges the Usage-Based Model of language, is in fact the grammar of a language.[3] The method is termed Multifactorial Usage-Feature Analysis (Glynn 2008, 2009, 2010, 2014e, 2014f; Krawczak and Kokorniak 2012; Fabiszak et al. 2014; Klavan 2014; Krawczak 2014a, 2014b; Krawczak and Glynn 2015, inter alios) or the Behavioural-Profile Approach (Gries 2006; Divjak 2006, 2010; Divjak and Gries 2006; Gries and Divjak 2009; Gries 2010, Gries and Otani 2010, Deshors and Gries 2014, Deshors 2014, Deshors forthc., inter alios) and has been applied to a wide range of phenomena.[4] However, for the reasons described above, adapting the method to the study of polysemy is not straightforward. The main limitation is that, in contrast to near-synonymy, there are no predetermined discrete categories whose profiles can be determined. In other words, in the description of near-synonymy, one can analyse two lexemes, morphemes, or constructions and examine their profiles to see how they are different or how they are similar. Since we wish to proceed with bottom-up, without positing lexical senses *a priori*, in polysemy research it is not clear what a behavioural profile would represent. In other words, how would we distinguish the "different" profiles associated with a single lexeme in such a way that we could argue to have captured the semasiological structure, viz. the polysemy network.[5]

[3] Geeraerts (2006), Glynn (2010, 2014a), and Gries (2010) consider, in greater depth, the implications of the Usage-Based Model for semantic research.

[4] The two labels for the method are effectively interchangeable. "Behavioural Profile" foregrounds the overall approach in theoretical terms, while "Multifactorial Feature" foregrounds the actual method of analysis. Three anthologies, Gries and Stefanowitsch (2006), Glynn and Fischer (2010), and Glynn and Robinson (2014), all largely devoted to the application of this methodology, are representative of the research tradition.

[5] This brief discussion omits other points of complexity. Perhaps the most important is the role of formal variation, especially at the constructional level, in semasiological structure. Simply put, how does one account for the relation between different senses and different formal contexts, such as part-of-speech or construction. This problem is the focus of Glynn (2015),

In response to this fundamental methodological hurdle, two lines of enquiry have thus far been taken within the tradition of Usage-Feature Analysis and the Behavioural-Profile Approach. A first line of enquiry is represented by Gries (2006) and Glynn (2014b), both of whom examine the polysemy of *run* by first identifying and annotating lexical senses in the sample. This is problematic for two reasons. First, an *a priori* choice of lexical senses leads to an analytical circularity which defeats one of the purposes of the Profile-Based Approach. If the method is to be used to test models of semasiological structure, proposed through Radial Set Analysis for example, then one cannot use those same senses to test the descriptive adequacy of the models. Second, lexical senses are necessarily discrete categories and, therefore, their use in semantic analysis predetermines a result constituted by discrete structure. Again this runs counter to the aims of the Behavioural-Profile Approach which seeks to identify non-discrete patterns in usage. Evidently, these patterns may or may not be interpreted as lexical senses but, in any case, the structure should "fall out" from the results of the analysis rather than be determined by choices made in that analysis.

A second approach, taken in Glynn (2009) examining *hassle*, Berez and Gries (2009) examining *get*, and Glynn (2010a) examining *bother*, is to add formal features to the semantic analysis.[6] These formal structures act as "anchors" in the semasiological variation identified in the feature analysis. Although some would argue that adding formal variation to the description of semasiological structure is crucial to research in polysemy, if the semantic features are structured relative to formal variants of a lexeme, then the analysis is of an onomasiological structure, not a semasiological one. In other words, if one identifies lexical senses associated with a specific formal context, then one is no longer looking at polysemy (the semantic variation associated with a single form) but at semantic variation between forms, or near-synonymy (see Glynn 2015 for more detailed discussion on this point.).

Within the Cognitive Linguistics paradigm, there are other methodological developments that seek to resolve the problems outlined above. One approach is to return to more traditional methods and develop better tests for establishing the functional–conceptual categories believed to explain semasiological

which concludes that, theoretically, neither word senses nor constructions exist; instead form–meaning pairing needs to be understood as many to many pairings of semantic and formal features.

[6] Gries (2006) also uses formal features to help determine semasiological structure. In effect, this study uses a combination of predetermined discrete senses and formal structures in the analysis of polysemy.

structure. Rastier (1987), Victorri (1997), and Tyler and Evans (2001) are representative of this line of research. Although improving heuristics for determining category membership is essential, for Multifactorial Usage-Feature Analysis just as for any manual semantic analysis, such an approach continues to assume that lexical senses can be discretely reified. Another approach is the application of Latent Semantic Analysis (Schütze 1998; Heylen et al. 2015; see Turney and Pantel 2010 for an overview). This computational method is related to the profile-based approach in that it examines context of use and seeks to identify semantic structure indirectly by identifying patterns in the use of a lexeme. The principal difference is that this approach examines extremely large samples, with high levels of dimensionality, but is restricted to purely observable characteristics of use (such as collocations and colligations), which are automatically analysed. Due to the large sample size, the results are arguably more representative than those obtained in Usage-Feature Analysis. However, it has yet to be demonstrated that observable features of use alone are sufficient to identify purely semantic structure. Moreover, since the analysis is entirely automated, despite a high level of objectivity, the accuracy of the analysis suffers due to misanalysis, or “noise.” The more detailed or rich the automated analysis becomes, the greater the quantity of noise. These two inherent limitations are yet to be overcome in latent approaches.

Presented in this article is a proof-of-principle study that seeks to show how manually annotated multidimensional clusterings of usage-features can be used as an operationalisation of lexical senses. Specifically, the method should enable the following:

- Quantification for inductive and repeatable analysis
- Identification of non-reified and non-discrete lexical senses
- Interpretation in terms of fuzzy set theory and prototype set theory.

Q2 Since Dirven et al. (1992), the method employed in this study has been systematically demonstrated to meet the first of these three criteria in onomasiological research. Generalising its inductive bottom-up approach to polysemy should be straightforward. The second criterion has yet to be demonstrated and the method’s ability to identify non-reified and non-discrete senses will be considered successful if the clusterings of usage-features come together in an interpretable and coherent manner. Obviously, the true success of this approach will only be possible to determine if the results here are confirmed using a different methodology, such as elicitation or experimentation in an independent study (cf. Klavan and Divjak, this issue, on the relationship between corpus and experimental methods of analysis). Finally, with regards to the third criterion, we will consider interpreting the results in terms of prototype effects and fuzzy membership.

2 Data and analysis: A case study in usage-based lexical polysemy

2.1 Data: A verbal lemma and personal online diaries

The lexical category under examination is *to annoy*. Derived forms, such as *annoyance*, *annoyingly*, *annoyedly*, and the adjectival uses were not included in the sample. Although the data extraction was designed to retrieve only verbal forms, some adjectival and gerundive uses were also returned. These were identified manually and omitted from the analyses. Other morpho-syntactic variation is left random in the sample, but coded for and reported below. In total, 500 occurrences were extracted with a considerable context (400 characters left and right) to enable Usage-Feature Analysis.

The sample was extracted from the *LiveJournal Corpus* (Speelman and Glynn 2005). This corpus is especially useful for Usage-Feature Analysis because it is sociolinguistically and stylistically homogenous. Manual semantic analysis requires relatively small samples, which, arguably, makes it impossible to claim language representativity. Instead of attempting to make claims about the structure of an entire language based on a small sample, focusing on a single text-type increases confidence in the representativity of the results, even if they are limited in scope. The underlying reasoning here is the empirical axiom that it is better to say more about less than less about more. Of course, this entails that in order to determine if the findings hold true for a broader cross-section of language, one needs to perform a repeat studies on different text-types.

The corpus consists of entirely online personal diaries. Although it is impossible to control for age, gender, and social class, the overwhelming majority of the authors are young people of secondary school or university age, blogging about their daily personal life. The corpus was compiled by crawling the *LiveJournal* blogging service, which is aimed at this particular socio-economic group. The sample of 500 occurrences was taken from British and American speakers, 50 % each. Dialect effects upon the semasiological structure are treated in Glynn (forthc.).

The lexeme *annoy* was chosen because it does not display a great deal of clear-cut semasiological variation. The reasoning here is to test the capability of the method in a situation where only subtle variation will be observed. As evidence for the semasiological simplicity, let us consider the lexicographic treatment of the lexeme. The *Oxford English Dictionary* defines the usage thus:

1. Verb intransitive: Be hateful, because of trouble. Archaic only
2. Verb transitive: Cause slight anger or mental stress
3. Verb transitive: Molest, injure, harass
4. Verb transitive: Damage something material. Dialectal only

Although in general definitions in the British *Oxford English Dictionary* and the American *Webster's Dictionary* tend to differ substantially, differences between the two definitions in this case are not remarkable. The definition offered by the American *Webster's Dictionary* is as follows:

1. Verb transitive: To disturb or irritate especially by repeated acts
2. Verb transitive: To harass especially by quick brief attacks
3. Verb intransitive: To cause annoyance

2.2 Analysis: Types of actors and usage-features

The general premise of Multifactorial Usage-Feature Analysis is to analyse characteristics of usage that may serve as indices of semantic structure. These features are traits or characteristics which are believed to come together to distinguish different language structures. They are sometimes referred to as "ID tags." Since we are considering a verbal category here, the obvious choice for the semantic analysis is the actors, or participants, involved in the event structure. In employing the actors and their relations, we are drawing upon Rudzka-Ostyn's (1988, 1989, 1995) early work in Usage-Feature Analysis in the description of communicative verbs. In the annoyance event, there are three possible actors: the Agent, the Cause, and the Patient. Each actor will represent a factor in the ensuing quantitative treatment of the results of the feature analysis. In this section, we consider the manual feature analysis of each of these actors.

2.2.1 Cause

We begin with the analysis of Cause, since it is the most problematic. Consider example (1):[7]

[7] All exampels are taken from the *LiveJournal Corpus* (Speelman and Glynn 2005).

(1) *I feel a tad immature and apologies if our hyperness annoyed you.*

In this example, the semantic frame and its instantiation are clear. The speaker is the Agent, his or her behaviour is the Cause, and the direct object is the Patient. However, in many instances, the categorisation of Agent and Cause needed to be conflated:

(2) *Dining tables annoy me.*

In example (2), no behaviour, event, action, or specific characteristic of the Agent is identifiable as responsible for the annoyance. The extended context (not included) reveals the speaker to be making a humorous remark about his or her preference for a computer chair. For this reason, we can assume that, at some level of analysis, it is likely that it is not the actual existence of dining tables that annoys the Patient but some specific characteristic of dining tables. However, we are not privy to what that actual Cause might be. In such instances, where the specific Cause is unknown, the Agent and Cause are conflated in the analysis. However, when an analytical distinction was possible, it was made.

(3) *ew … my font annoys me.*

Although it could be argued that in example (3) the Agent and Cause need to be conflated, as in (2), the use of the exclamation *ew*, which denotes disgust or distaste, gives a sufficiently clear index of the actual Cause of the annoyance. Indeed, it can be inferred that it is the aesthetics of the font that causes the response rather than the font itself; such an example would then be annotated as Cause – "Aesthetics."

The feature categories used for the analysis of the actor Cause are partially taxonomic. At first, a highly detailed (fine-grained) set of causes was annotated. From this, certain highly frequent fine-grained distinctions were retained, while less frequent features were collapsed into semantically more abstract categories. The decision to retain these fine-grained features and proceed with a taxonomically complex feature set is hoped to reveal the more subtle variations associated with the lexeme. The four coarse-grained or high-level categories include "Concrete Thing," "Specified Activity," "Non-Specified Activity," and "Specified SoA" (SoA = state-of-affairs). The category of "Concrete Thing" was exemplified above in (2) but is repeated here as (3a). Examples (3b)–(3d) are typical of the other high-level, or schematic, features.

(3) a. Cause – 'Concrete Thing'
Dining tables annoy me.
b. Cause – 'Specified Activity'
...his gasps continued annoying Draco.
c. Cause – 'Non-Specified Activity'
...he also managed to annoy several others who I won't mention.
d. Cause – 'Specified State-of-Affairs' (SoA)
Yet my skateboard is scratched and chipped and it annoys me.

With regard to the features exemplified in (3b) and (3c), the label activity is taken to include not only activities, but also actions, events, and behaviour that could not be clearly described as one of the more specific types of activities–actions–events described below. The distinction between specified and non-specified could also be rephrased as overt and covert. Even with considerable context, it is often not possible to ascertain the actual nature of the activity and in these situations the category "Non-Specified" was used. Example (3d) is an instance of a Cause that can be described as a state-of-affairs. The term originates in Pragmatics but is not employed here in its strict reading. Generally, when the Cause was some status in the world, this category was ascribed.

The more specific Cause features include six categories which were found to be particularly frequent with respect to the use of the lexeme *annoy*. Examples (4a)–(4f) represent the uses in question.

(4) a. Cause – 'Emotional Pain'
i want to be his firend not just some girl he knows i want to knwo everything about him i want to be his favitore person in the whole world i want to have convos that we shouldnt be talking about i want his trust but another person has it and that person annoys the fuck outta me just cuz im jealous of that person
b. Cause – 'Thoughts'
Fear controls me all the time and it annoys me the most of all things.
c. Cause – 'Imposition'
guys that like you who constantly annoy you thinking it might work...?
d. Cause – 'Interruption'
4 AM we called Cathy up just to annoy her.
e. Cause – 'Repetition
...hospital was much more boring this. And I was on a drip. I'm not going to go into details of the stay but it was ... tiring and annoying and blech.
f. Cause – 'Aesthetics'
ew ... my font annoys me.

In (4a), the use labelled with the feature of “Emotional Pain” should be quite clear. It categorises any use where the Cause is in the mind of the Patient but, importantly, where some painful or emotionally detrimental experience is at stake. The feature Cause – “Thoughts,” exemplified in (4b), is quite close to “Emotional Pain” (4a), saves that the Cause here is more intellectual and less associated with emotional suffering. Despite the similarity and possible overlap between these uses, most occurrences clearly belonged to one or the other. The feature of “Imposition” was annotated when the volition or intention of the Agent takes precedence over that of Patient; against the Patient’s wishes (4c). “Interruption” as a Cause of annoyance should be self-explanatory (4d), and in (4e) we see an example of a Cause which is some event, activity, action, or state-of-affairs that bores the Patient. Example (3), repeated here as (4f), is a typical example of Cause – “Aesthetics.”

2.2.2 Patient

The actor, or factor, of Patient was the most straightforward. Due to the lexical semantics of the verb in question, all the Patients were human and, due to the “diary” nature of the text, most often in the first person. Three usage-features were identified: “1st Person,” “Not 1st Person Specified Human,” and “Not 1st Person Non-Specified Human.” The general category of “Not 1st Person” is used because there are so few second person examples and when they occurred, it was far from clear that it was actually a second-person referent and not third-person generic *you*. By “Specified Human” is meant a person known to the speaker rather than a generic human referent. The distinction, therefore, in the “Not 1st Person” category is effectively between generic and specific.

(5) a. Patient – ‘Human Specified 1st Person’
It annoys me when standards interrupt my personal way of thinking.
b. Patient – ‘Human Specified Not 1st Person’
And I won’t have to annoy the people I love.
c. Patient – ‘Human Non-Specified Not 1st Person’
You know how to have the best time playing, and can annoy adults to no end.

2.2.3 Agent

This factor is more difficult to break down into constituent features. “Specified” and “Non-Specified” humans were identified in the same manner as for Patient.

Three other categories were added: “Event-activity,” “States-of-Affairs,” and “Inanimate Thing.” The problem is that semantically there is effectively a continuum from “Specific Event,” to “Activity,” to “State-of-Affairs,” to “Abstract Thing,” and to “Concrete Thing.” Although most uses where immediately categorisable as one of the other, some lay at the border between these categories. As a general rule, “Event-Activity” was annotated for verbal profilings, “Inanimate Things” for nominal profilings, regardless of how abstract the referent was, and “State-of-Affairs” for anaphoric references to complex propositions. Although using part-of-speech as a guide seems reasonably categorical, the reality of the data, even with considerable context, means that often it was not clear what the referent was. This is an inherent limitation of the method and needs to be explicitly signalled as such.

(6) a. Agent – ‘Event-Activity’
...so it annoys me when people come up to me and expect help.
b. Agent – ‘Inanimate Thing’
so we have like no programes on here at the moment, no msn messanger but im not in a great hurry to install that back on, it sometimes annoys me;p
c. Agent – ‘State-of-Affairs’
It really annoys me when employers assume you don’t know wwhat you’re rights are.

2.2.4 Inter-rater agreement

The factor, or semantic category, of Cause is particularly difficult to operationalise, the different features in question being both taxonomically related and not discretely distinguishable. This point remains the single most important weakness in this method. Future research needs to prioritise both improved heuristics for operationalisation and multiple coders, combined with more rigorous statistical measures to ascertain reliability of manual semantic annotation (see Artstein and Poesio 2007 for an overview of the issue).

A subsample of the Cause category/factor was submitted to a secondary analysis in order to ascertain the reliability of the subjective analysis. The second annotator was not trained through corrections on a preliminary subsample. Instead, the annotator was given only the description offered in Section 2.2.1. The subsample, consisting of 80 examples, was determined using power analysis, based on a confidence level of 95 % and a confidence interval of 10, for the population of the entire sample of 500 occurrences. The unweighted Cohen’s

Kappa achieved 0.766, with a *z*-score of 8.64. Given that the analysis was performed without prior training and that the taxonomic relations between many of the categories are not discretely distinguishable, this score, while not extremely high, is adequate to show that the analysis is reliable.

3 Results – Mapping semasiological variation and structure

3.1 Formal variation and semantic distribution

The aim of the present study is to demonstrate proof-of-principle that multifactorial Usage-Feature Analysis is able to identify a semasiological structure in a way that is readily falsifiable and without positing reified discrete lexical senses. Importantly, we wish to demonstrate the descriptive power of the method in situations where it is not immediately obvious that semantic distinctions exist. Previous research in polysemy using this method (Gries 2006; Glynn 2010, 2014a) has combined formal and semantic features in the analysis. In these previous studies, formal distinctions that closely mirrored semantic distinctions improved the descriptive power of the results, but some of the form–sense distinctions identified were relatively clear-cut. For this reason, it is important in this study that the distributions of the semantic features do not reveal any immediately obvious and, therefore trivial, structures. The main concern in this regard is the redundancy between certain Agents and Causes. The effects of this redundancy are discussed below. In light of this, let us consider the formal and semantic distribution of the observed data. Table 1 summarises the different parts-of-speech and morpho-syntactic constructions found to be associated with the lexical category.

Table 1: Morpho-syntactic variation of *annoy*.

Form	Freq.	Form	Freq.	Form	Freq.
Adjective attributive	5	Transitive Cx	415	Trans. ablative (*about*) Cx	1
Adjective predicative	7	Trans. instrumental (*with*) Cx	8	Trans. ablative (*because*) Cx	1
Gerund	2	Resultative Cx	1	Trans. ablative (*for*) Cx	1
the X *out of* Y Cx	55	Intransitive Cx	6	Trans. ablative (*with*) Cx	2

Of the 415 Transitive uses, 252 were in the simple present, 111 were in the present or past continuous, 60 were infinitival, and 3 modal. A comparable distribution between simple, continuous, and non-finite was found in all of the other verbal profilings. Even these simple findings show the inadequacy of the dictionary definitions. Although the adjectival uses only represent a small minority of the uses, given that the sample is only 500 occurrences, the 12 adjectival instances are not negligible. These occurrences and the gerundive uses are excluded from the analysis below: Their use is systematically distinct, in that they have no overtly expressed patient. Since features characterising the Patient constitute one of the main factors in the analysis, the discrete behaviour of these examples, with regard to the characteristics of the Patient, would obscure other less clear, but more insightful patterns in the use of the examples.

Table 2 summarises the distribution of the usage-features. We see that although the factors Cause –"Aesthetics" and Cause – "Repetition" are both relatively infrequent, the majority of the features are widely distributed.

Table 2: Variation of semantic actor types of *annoy*.

Cause	Freq.	Cause	Freq.	Agent	Freq.	Patient	Freq.
Concrete thing	25	Emotional pain	38	Event	55	HumSp 1stPrs	296
Specified activity	92	Imposition	34	HumNSp	56	HumSp N1stPrs	152
Non-specified activity	152	Interruption	76	HumSp	237	HumNSp N1stPrs	42
Specified SoA	35	Repetition	7	SoA	83		
Aesthetics	9	Thoughts	33	Thing	70		

3.2 Semasiological usage-patterns

Having summarised the observed variation, let us now turn to identifying structure in that variation. Figure 1 contains the visualisation of a multiple correspondence analysis. This is a multivariate dimension reduction technique that is designed to identify correlations in complex data. The results of the annotation (usage-feature analysis) of the factors Agent, Patient, and Cause are treated mathematically in order to identify marked correlations and anti-correlations, as presented in Table 2. The aim is to reveal underlying structures in the form of associations in use. Indeed, these associations are understood to constitute the usage-patterns of polysemy. The representation is based on a chi-square distance measure (weighted Euclidean distance matrix), the results of which are visualised in two dimensions.

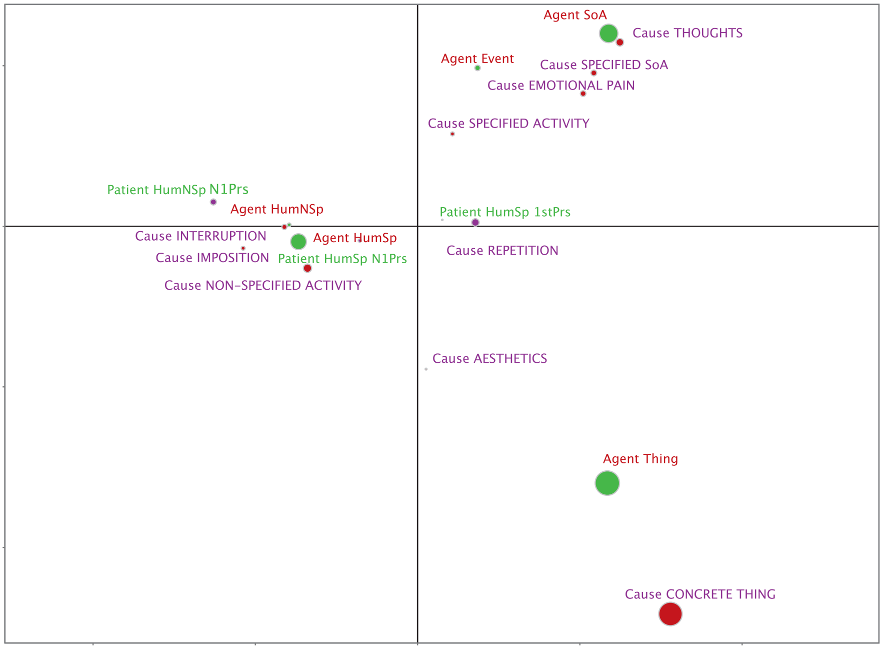

Figure 1: Agent, patient, and cause correlations for *annoy*.
Multiple Correspondence Analysis (inertia 69.7 %)

In Figure 1, proximity between the data points represents degree of association. The size of a data point represents the contribution of that feature to the overall structure, that is, relatively, how much it is responsible for the associations revealed. Data points that lie further away from one of the two axes are more distinctive than data points that lie closer to one of the axes. This means that features close to the centre (the intersection of both axes) are indicative of more general, non-distinctive usage. In this light, clusters of features that are central could be argued to be a frequency-based prototype. This point will be raised in Section 4 (see Glynn 2014d for a more detailed explanation on the functioning and interpretation of correspondence analysis).

This relatively simple plot appears to divide into three vague clusters of usage-features. These three clusters are dominated by the Agent types: In the top-right quadrant, we see “State-of-affairs” and “Event” coming together; close to the centre on the left, we see “Specified Human” and “Non-Specified Human;” and finally, in the bottom right quadrant, the Agent type “Thing.” This third “cluster” is made up of the features Agent – “Thing” and Cause – “Thing,” which is not an informative result except that it clearly represents a distinctive use. The feature Cause – “Aesthetics” lies between these two data points and the other clusters. This means that it is highly associated with Agent – “Thing,” but not distinctly so. Nevertheless, the profile of this bottom-right cluster is not particularly informative in terms of semantic profiles.

The cluster of features in the top-right quadrant offers a more informative profile. That the two semantically similar Causes, namely, “Thoughts” and “Emotional Pain” come together is a sign of semantic structure. When the quantitative analysis reveals that two closely related semantic features correlate (behave similarly), we can first interpret this as an indicator that the given distinction is not needed in this instance. However, such a correlation also suggests that the quantitative analysis is automatically identifying underlying semantic relations. The two other Causes, “SoA” (state-of-affairs) and “Specified Activity,” represent a redundant correlation with the Agents “Event” and “SoA.” Nevertheless, if we compare these Agents and Causes in this cluster with the other two clusters, the structure becomes apparent. In the cluster at the top-right, the Agents are inanimate, but not things. In other words, the Agents are processes in the world, but the actual cause of the experience is internal to the Patient, his or her thoughts or emotions. This profile is distinctive and coherent, especially in contrast with the other clusters. The left-hand cluster is characterised by people annoying people by doing things, often specifically interrupting or imposing. This usage would probably fit the stereotypical scenario associated with the lexical category. Whether it would prove to be prototypical in terms of how speakers conceptualise the lexical category is an empirical question. However, in terms of frequency-based data, this cluster lies near the centre of the plot, which means it is the least distinctive in usage and therefore, arguably, the most typical of the usage clusters.

Given the nature of the data, the role of Patient is expected to be less informative. However, the correlations observed draw a picture that is intuitively sound, but with one important surprise. The data point Patient – “Human Specified 1st Person” is equidistant between the three clusters and near the intersection of the two axes, both correlation positions indicating that it is non-distinctive in use. In other words, it is neutral with regards to the semasiological variation. Since the texts are made up of personal descriptions of daily life, this is hardly surprising, but its placement helps to demonstrate the reliability of the dimension reduction technique. More informative and slightly surprising is that the cluster on the centre-left, which we can assume is a representative of the least distinctive cluster, is relatively associated with “Non-1st Person” Patients, regardless of whether they are generic or not. The two data points Patient – “Human Non-Specified Not 1st Person” and Patient – “Human Specified Not 1st Person” are central to the cluster (note that the actual data point for “Human Specified

Not 1st Person" is partially hidden beneath the label Agent – "Human Specified"). This clustering suggests that although the first person is typical of all uses, the cluster is markedly associated with uses where the speaker is referring to Patients other than him- or herself (at least relative to the other usage-clusters). Given the personal nature of the texts used, this is noteworthy.

It must be stressed that correspondence analysis is an exploratory technique and so cannot be used to make statements about the distribution/structure beyond the sample. In other words, the only way to make claims about language structure, here the polysemy of the lexeme *annoy* in English, is to repeat the analysis with a second sample. There are statistical means for achieving these goals, but they are not the focus of the current study. However, we can determine how well the two-dimensional plot captures the complexity found in this specific sample. Tables 3 and 4 present the numeric output of the correspondence analysis presented in Figure 1.

Table 3: Scree plot of Multiple Correspondence Analysis (Greenacre adjusted).

Dim.	Eigen value	Inertia	Cum. Inertia	Scree
1	0.158857	48.8	48.8	*************************
2	0.067994	20.9	69.7	***********
3	0.022388	6.9	76.7	****
4	0.009902	3.0	79.7	**
5	0.004173	1.3	81.0	*
6	0.000241	0.1	81.0	
7	00000000	0.0	81.0	
8	00000000	0.0	81.0	

Table 3 presents a breakdown of the explained variance, or inertia, represented in the correspondence plot above. Mathematically, the analysis has 17 dimensions, something impossible to conceive for the human mind. In order to understand the relations in such a high-dimensional space, correspondence analysis attempts to collapse this complexity to just two dimensions, which it then visualises. These scores tell us how well the plot visualises the structure of the data. The plot is two-dimensional: distances, portrayed along the *x*-axis (from right to left), represent 48.8 % of the structure; adding the *y*-axis (top to bottom) adds 20.9 % explanation to that, making a total of almost 70 % accurate representation across two dimensions. Given the complexity of the analysis, this is a high score, and the plot can be said to be a stable representation of the observed variation. Importantly, it should be noted that adding a third dimension would

Table 4: Numeric output of Multiple Correspondence Analysis (Greenacre adjusted).

Col.	Feature	Dim. 1	Contr.	Dim. 2	Contr.	Mass	Quality
1	Cause AESTHETICS	27	0	326	9	6	420
2	Cause CONCRETE THING	803	109	884	310	27	774
3	Cause EMOTIONAL PAIN	525	42	−303	33	24	810
4	Cause IMPOSITION	−554	41	50	1	21	648
5	Cause INTERRUPTION	−423	55	1	0	49	699
6	Cause NON-SPEC. ACTIVITY	−349	76	95	13	99	604
7	Cause REPETITION	78	0	−15	0	5	50
8	Cause SPEC. ACTIVITY	111	5	−211	39	59	287
9	Cause SPEC. SoA	559	41	−350	38	21	593
10	Cause THOUGHTS	642	56	−420	56	22	792
11	Patient HumNSp 1stPrs	−648	75	−55	1	28	835
12	Patient HumSp 1stPrs	183	43	−9	0	202	982
13	Patient HumSp N1Prs	−184	22	33	2	103	581
14	Agent Event	190	8	−361	67	35	339
15	Agent HumNSp	−407	36	−3	0	35	498
16	Agent HumSp	−378	139	35	3	155	780
17	Agent SoA	606	128	−440	157	55	746
18	Agent Thing	602	123	585	272	54	778

only contribute a further 7 %. This means that, structurally, the complexity of the data can be adequately represented in two dimensions.

Table 4 offers further information about the reliability of the representation of the results. Each of the features is listed along with their relative contribution (Contr.) to the two axes (Dim. 1 and Dim. 2). The Mass score indicates the amount of weighting the analysis has had to give to each feature in order to account for relative frequencies and, finally, Quality is a statistic that permits us to determine the accuracy of representation of a specific data point. This statistic was developed by Greenacre (2007) and although there is no simple benchmark for reliability, obviously the representation of Cause – “Repetition” is extremely low and any interpretation as to its associations in the analysis should be treated with great caution. Lastly, it should be noted that the analysis itself was performed on a Burt matrix using Greenacre’s (2007) adjusted algorithm.

3.3 Semasiological usage-structure

Having identified the correlations and their reliability, we can attempt to ascertain underlying structure. In order to do this, we will employ two techniques. First, we perform a hierarchical cluster analysis on the results of the multiple correspondence. Second, we employ a partition cluster analysis directly upon the configurations of features.

The two plots in Figure 2 show how, mathematically, the dispersion in the correspondence analysis can be clustered. The degree of overlap between each of the three colours represents the difficulty the analysis has in distinguishing the groupings discretely. Although we do not expect discretely distinguished clusters (indeed, on the contrary), a high degree of overlap would suggest that the proposed underlying structure of a three-way distinction cannot account for the behaviour of the data. The cluster analysis produces a scree plot of the inertia for different numbers of clusters, and three clusters were found to be the most distinct "elbow," that is where the best balance between parsimony (or simplicity) and inertia (or explanation) is found. In the bottom-left quadrant of the factor map on the right, we see the hierarchical cluster analysis in two dimensions with the "cut" at three clusters. Further subdivisions appear plausible, and we will return to this point below.

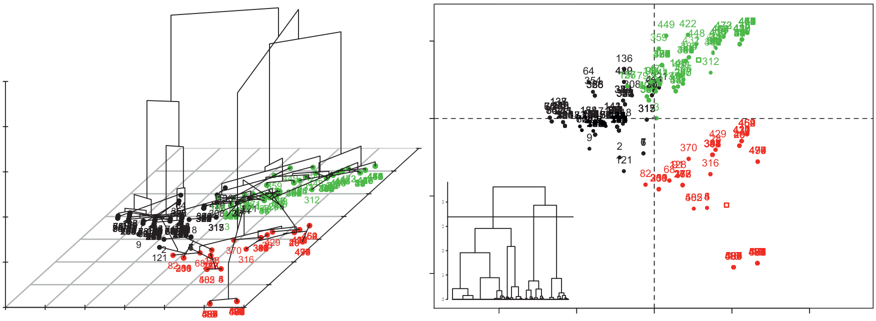

Figure 2: Agent, patient, and cause agglomerative clusters for *annoy*.
Hierarchical cluster analysis on correspondences between three actors

Although the three-dimensional representation of the clustering of the examples on the right can be difficult to interpret, it should be clear how the hierarchical clustering proceeds and identifies three groups. The plot on the right merely colours each of the examples as belonging to one of the three clusters. The Agent/Cause – "Thing" cluster in the bottom right is totally distinct, with only some examples lying between it and the left-hand cluster. There are no examples lying between it and the top right-hand cluster. In contrast, the two top clusters gradually merge into each other suggesting a semantic continuum. However, importantly, these clusters are clearly distinguishable with only a small amount of actual overlap. Note how in the two-dimensional dendrogram in the bottom left-quadrant of the factor map, the left-hand cluster is internally very complex with many sub-clusters. This internally complex cluster could be argued to consist of two sub-clusters. If we compare these results with the correspondence analysis in Figure 1, we see that this complex cluster is made up of Agent – “States-of-Affairs” and Agent – “Events,” which are correlated with Cause – “Thoughts” and Cause – “Emotions” but also redundantly correlated with “States-of-Affairs” and “Activity.” It is entirely possible that this cluster is representative not of a single semasiological structure but, in fact, two.

In Figure 3, the actual observations (as opposed to the results of the correspondence analysis) are clustered using a partition cluster analysis, based on medoids. The *k*-medoid algorithm is essentially the same as the more widely used *k*-means method, except it is less sensitive to effects of outliers. In a *k*-medoid partition cluster analysis, the number of clusters is predetermined, and the analysis attempts to “sort” the examples accordingly. In the results, the more distinct the resulting clusters, the more confident one is that those clusters represent an underlying structure in the data. A series of cluster analyses were performed, beginning with two clusters ($k = 2$) and subsequently adding one cluster until $k = 6$. A clustering of $k = 3$ was the best fit of the data, and the two plots in Figure 3 are based on this three-way division of the examples.

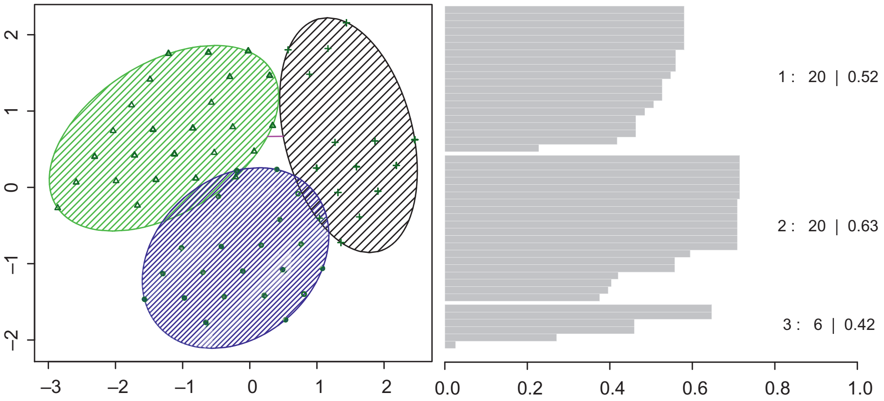

Figure 3: Agent, patient, and cause medoid clusters for *annoy*. Partition cluster analysis of the three actors

On the left, a clustering plot of unique data points (feature configurations) was produced using the Euclidean distance matrix, which is the simplest matrix. The plot explains 73.2% of the variability (principal components on the correlation matrix). The three clusters are represented by different symbols, and ellipses are drawn around each group. Although there is some overlap between the clusters, their overall distinctiveness is clear, and only the bottom cluster reveals substantial overlap. However, in the right-hand cluster, it is clear that there are, in fact, two subgroupings, with some of the feature configurations positioned at the top, closer to the top-left cluster, and others positioned towards the bottom and closer to the bottom-centre cluster. More detail on the use and interpretation can be found in Horvath (2011). Despite the apparent clarity of the clustering, the silhouette plot on the right reveals weaknesses.

The partition cluster in the silhouette plot used the Manhattan distance matrix; like the Euclidean distance matrix, this is an unweighted and quite "neutral" dissimilarity measure. Priness et al. (2007) and Divjak and Fieller (2014) offer discussion of the different distance matrices. In the silhouette plot, the width of each "flag" represents the quality of the cluster. Configurations represented by lines that drop to the left of the "flagpole" (the left-hand vertical line of the plot) have been misclassified. The first thing to note in the silhouette plot above is that there are no misclassifications (lines to the left of the "flagpole"), so we can infer that the clustering does not force examples into the groups where they should not be. However, we know from the cluster plot on the left that the problem is the internal coherence of the clusters. UNESCO (2013) offers the following breakdown of average silhouette scores as a guide – <0.25: no substantial structure is found; 0.26–0.50: structure is found but it is weak; 0.51–0.71: a reasonable structure is identified; >0.71: a strong structure is identified. The average silhouette or flag width is 55, and the third cluster is only 46. This does not mean that the quality of the clusters is poor, only that the picture is far from obvious, and the internal structure of the clusters is only weakly coherent. Claude (2008) offers more detail on the interpretation of such methods. For our purposes, however, we can safely say that there exists a three-way underlying structure in the behaviour of the actor features for the use of *annoy*, but that these three clusters of uses are only loosely coherent, the internal structure of each cluster being relatively heterogeneous. Such a quantified interpretation of the usage associated with the lexeme *annoy* could reasonably be argued to represent semasiological structure in non-reified and falsifiable terms.

If we take this three-way clustering of the results, we can summarise the configurations of features found in the correspondence analysis:

Cluster 1: Cause – 'Thought', 'Emotional Pain'; Agent – 'Event', 'SoA'; Patient – '1st Person'
Cluster 2: Cause – 'Interruption', 'Imposition', 'Non-Specified Activity; Agent – 'Human'; Patient – 'Not 1st Person'
Cluster 3: Cause – 'Aesthetics', Thing'; Agent – 'Thing'; Patient '1st Person'

3.4 Confirming usage-patterns

Having identified an underlying semasiological structure and identified clusters of usage-features that can be understood as operationalisations of non-reified non-discrete lexical senses, we can now attempt to calculate the significance of the associations that are posited to represent these senses. The manual analysis of the data in this study means that, relative to the complexity of the analysis, the sample is too small to perform predictive modelling. However, based on the results of the correspondence analysis, we can perform a loglinear analysis if we simplify the feature analysis by "clumping together" some of the more fine-grained distinctions. This technique will identify significant correlations and anti-correlations between the usage-features. In other words, it will produce a quantified statement about the representativity of the correlations (which structure the patterns) observed in the above statistical analyses.

Figure 4 is mosaic plot visualising the results of a loglinear analysis of the three factors: Agent, Patient, and Cause. Due to the small sample size, no interactions were included in the analysis, and the Cause categories of

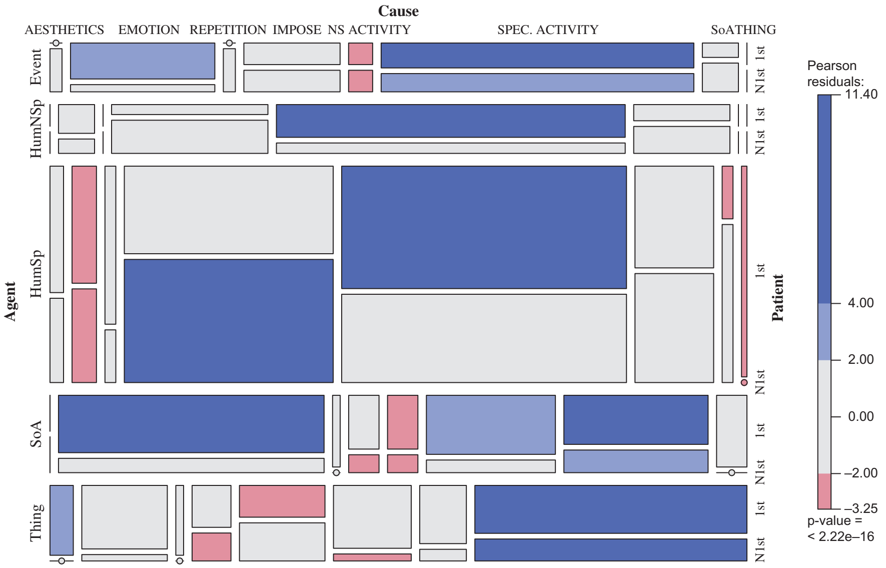

Figure 4: Three correlations between Agent, Patient, and Cause
Mosaic visualisation of loglinear analysis

“Interruption” and “Imposition” were conflated and renamed as “Impose.” Similarly, the categories of “Emotional Pain” and “Thoughts” were conflated to a single category “Emotion.” Finally, for the category Patient, “Non-Specified Human Not 1st Person” and “Specified Human Not 1st Person” were conflated to simply “Not 1st person” (N1st). Although such simplification is regrettable, it was necessary in order to apply loglinear analysis. The category conflation was based on the observations in Figure 1 and only features that appeared to be highly associated as well as semantically related were conflated.

Mosaic plots are notoriously difficult to interpret, but this is a result of the complex information they represent. Figure 4 represents the interaction of three factors: Agent, Cause, and Patient. Each of these factors has several levels, or linguistic usage-features. The loglinear analysis examines every possible combination of the features and determines whether that co-occurrence of features is significantly higher or lower than one would expect if that correlation were simply the natural arbitrary variation of a sample of uses. In other words, significant correlation and anti-correlation can be interpreted as probably representative of the population; in the case at hand, the language of personal online diaries. Lack of significance, however, does not necessarily mean that the results identified in the correspondence analysis were wrong or merely the result of chance; it could also mean that there are insufficient data to be confident that they are representative. The plot should be read from left to right, each row a set of boxes, and each box representing a three-way correlation between the features of the three factors: Agent, Cause, and Patient. The size of the box represents the frequency of the three-way co-occurrence. Grey represents non-significant correlation, blue represents significant correlation, and red represents significant anti-correlation; the darker the shade the stronger the (anti-)correlation found.

Some of the correlations are redundant, in that they are a result of underspecified Causes, as explained in Section 2.2. For example, the dark blue boxes in the bottom-right reveal that Agent – “Thing” and Cause – “Thing” are highly correlated. Similarly, just above these boxes, in the second row from the bottom, Agent – “SoA” and Cause – “SoA” are found to be significantly associated. Such trivial correlations are, obviously, not informative. However, the fact that Agent – “SoA,” Cause – “Emotion,” and Patient – “1st Person” are highly and significantly correlated is informative and supports the correlations summarized as cluster 1 above. The anti-correlation that Agent – “Human Specified” is not associated with Cause – “Emotion” is important to show how distinctive cluster 1 is from cluster 2. The correlation between Agent – “Human Non-Specified” +

Cause – “Specified activity” + Patient – “1st Person” in the second row from the top also confirms this clustering.

In the centre of the plot, two important correlations are identified as distinctions for Agent – “Human Specified,” which is highly associated with Cause – “Impose” if the Patient is not the speaker and highly correlated with Cause – “Non-Specified Activity” if the Patient is the speaker. This is direct confirmation of cluster 2. Cause – “SoA” and Cause – “Thing” are statistically disassociated from Agent – “Human Specified” when the Patient is the “1st person.” This is a result of the distinctiveness of cluster 3.

Table 5 includes a summary of the significant correlations revealed in the loglinear analysis. The corresponding cluster is also listed. One three-way correlation identified in the correspondence analysis above was not confirmed: Agent – “Human Specified” + Patient – “Non 1st Person” + Cause – “Non-Specified Activity” for cluster 2. However, the failure to find significant correlation here does not undermine the overall profile of the cluster.

Table 5: Significant correlations relative to three-way clustering of features.

Agent		Patient		Cause	Cluster
A-SoA	+	P-1stPrs	+	C-emotion-thought	Cl. 1
A-event	+	P-1stPrs	+	C-emotion-thought	Cl. 1
A-human specified	+	P-N1stPrs	+	C-impose-interrupt	Cl. 2
A-human non-specified	+	P-1stPrs	+	C-non-specified activity	Cl. 2
A-human specified	+	P-1stPrs	+	C-non-specified activity	Cl. 2
A-thing	+	P-1stPrs	+	C-aesthetics	Cl. 3

Having identified fine-grained usage patterns, held to be indicative of semasiological variation (Section 3.2) as well as having determined underlying structure in those patterns (Section 3.3) and statistical significance for core correlations in those patterns (Section 3.4), we could now create a set of configurations for each of the “senses” posited. These configurations could be formalised in a manner similar to Brugman (1983)/Lakoff (1987), with each of the actors (Agent, Patient, and Cause) representing a dimension and overlap in correlations representing “links” in the radial network. Representing the results in the box- representation proposed by Geeraerts (1995) and commonly used in prototype studies would also be straightforward. However, these options would defeat one of the aims of the study, namely, develop a means for the non-reified description of semasiological structure.

4 Discussion: Frequency-based and multidimensional set structure

The first two aims of this study have been fulfilled. The first aim was to demonstrate that multifactorial Usage-Feature Analysis could quantify lexical semantic structure in a way to produce empirical and inductive results along with the possibility of repeat analysis. Although predictive confirmatory modelling has not been performed, the statistical significance of key correlations has been established, despite the relatively complex nature of the results. Nonetheless, future work will be needed to extend quantitative methods to predictive modelling. Regardless of the quantitative means for determining the accuracy of the results, due to the observational nature of the data, the results themselves can be easily falsified by repeat analysis – taking a second sample and re-applying the same analysis and comparing the results.

The second aim was to demonstrate that usage-features could be used to indentify non-reified lexical senses. At this point, we need to be more careful. Obviously, since the semasiological structure of a lexical category is not known *a priori*, it is not possible to determine whether that semasiological description is accurate. In order to determine descriptive accuracy, evidence from other sources would be needed to corroborate the findings. For this reason, the aim here was not one of descriptive accuracy but of descriptive method. In Section 1, it was pointed out that a description of lexical semantic structure in terms of non-discrete non-reified lexical senses would be expected, given Cognitive Linguistic theory. Despite this, previous research in the field has treated such categories/senses as “nodes” in a network or as sets of features in certain configurations. The problem lies in the categorisation of individual uses into these categories. The bottom-up approach employed in the present study, combined with the simple assumption that the structures we are seeking to identify need not be discrete, appears to have been successful. Although the semasiological structure indentified is reasonably simple, a three-way distinction of meaning roughly divided by various configurations of actor features, it is intuitively reasonable. Most importantly, the Usage-Feature Analysis does produce a semasiological description where sense structure is identified as relative clusterings of features based on the relative co-occurrence of features of instantiations, or uses, rather than instantiations matched to categories predetermined by feature configurations.

The third goal of the study was to develop an empirical quantified method for description of semasiological structure in terms of contemporary set theory.

This is surely the most difficult aim to achieve. Fuzzy set theory allows us to quantify specific dimensions as continua but still permits those dimensions to contribute to structuring. For example, drawing on the assumptions of fuzzy set theory, Cause – "Emotional Pain" and Cause – "Thought" could be treated as two non-discrete categories, perhaps lying on a continuum. Since the feature analysis necessitates discrete labels, at the analysis stage, the method fails to apply the principle of fuzzy categorisation. Although the distance matrix used to produce the correspondence analysis converts the relative frequency of these categories to a continuous scale, the labels themselves remain discrete. The plots do represent these labels in an analogue manner, but this does not change the fact that the actual analysis is categorical. Glynn and Krawczak (2014) have tested the possibility of employing 9-point Likert scales in Usage-Feature Analysis with some success. However, the factors under analysis, there were bipolar, along a single dimension, such as positive and negative evaluation. It is not immediately obvious how such a heuristic could be applied to the complex feature set associated with actors in an event scenario, such as that examined in the present study.

Turning to prototype theory, it would seem that method enjoys more success. The crucial problem is that the individual "senses" need to be described as part of a prototype-based structure relative to the lexeme, but in turn, the instances of use must be described in terms of a prototype-based structure relative to the senses. By working entirely bottom up and treating senses as simply complex clusterings of features of instances, this problem is resolved. As a result, we can safely say that generality of the use of features (i.e. the extent to which they are non-distinctive across the clusters) is an operationalisation of prototypicality both at the lexical "sense" (type) level and the individual instantiation (token) level.[8]

[8] An interesting aside that needs to be made at this stage concerns Langacker's (1987) Schematic Network Model of categorisation and polysemy. His work was crucial in early network theory. Just like Lakoff, Langacker used metaphors of nodes to talk about senses, and although he goes further to stress the negotiated usage-based nature of these senses, he is explicit about their reified and discrete nature: "each node in a lexical network represents a different established usage; in combination with the phonological pole, it defines a distinct semantic variant of the lexical item" (Langacker 1987: 384). However, Langacker speaks of elaboration and instantiation of (sub)schemata. Some of his discussion appears to be a usage-based re-interpretation of the Structuralist work on the *Gesamtbedeutung*, or the "aggregate meaning" of a semantic category. This reading of his work is reinforced by his use of phonemes and allophones to explain categorisation. This idea could be operationalised using the method proposed in this study by treating the features that are shared by all uses of a lexeme as constituting the *Gesamtbedeutung*. Although this is not the same as the instantiation of an abstract schema producing a subschema, in Langacker's understanding, it could be a

Two essential caveats are warranted in this regard. First, this operationalisation assumes that (relative) frequency can be used to identify conceptual prototype effects. We know *a priori* that this is not entirely true. What is most typical in a system is not necessarily the most important in its structuring. In many situations, frequency and importance overlap, but often they do not. Although the effect of frequency on language acquisition (and, therefore, on an individual's competence and, by extension, that of a language community) is beyond question, the role of prominence is also essential. Put simply, not every input (instance/instantiation/token) in the system is equal in its impact on the structuring of that system. Extrapolating from this premise, it becomes obvious that a combination of prominence and frequency both play a role in producing conceptual structure, and by extension, language structure. If these deductions are true, then frequency on its own is an imprecise operationalisation of conceptual structure. This does not necessarily mean it is inaccurate but it means that one must be cautious of frequency as an unique index of semasiological (conceptual-lexical) structure. Ideally, an operationalisation of conceptual structure should account for both typicality and prominence simultaneously. It has been argued elsewhere that frequency cannot be used to operationalise prominence, only typicality (Glynn 2010b, 2014c). Until an operationalisation that accounts for both structures of typicality and prominence is developed, frequency/typicality is an extremely useful (if incomplete) index for prototype effects.

Second, an inherent problem with observational data of this kind is the inability to account for what we may call semasiological salience. Geeraerts (1993b) examines onomasiological salience, which can be understood as the relationship between designata and denotata, or as the frequency (in terms of ratio) that a given lexeme is used to refer to a given referent relative to other parasynonyms that may also be used. Ideally, a similar ratio is needed in semasiological research. Leaving aside the difficulty of extending this premise to lexical senses (reified or not), it is difficult to determine the frequency of features relative to other uses beyond those associated with the lexeme under investigation. Without this information, the contribution of a given feature to the prototypicality of the sense cluster to which it contributes is not possible to ascertain. An example should clarify the problem. In the analysis above, in terms of the relative frequency of the different Patient types, "1st Person"

reasonable operationalisation. The problem, of course, will be the same as the earlier work on *Gesamtbedeutungen* – what happens when there are no features shared by all the uses? As M. Fabiszak points out (p.c.), this situation could be explained in terms of Wittgenstein's idea of family resemblance. Future research will have to consider these possibilities.

could be argued to be a prototypical feature, or at least a feature that contributes to typical usage. This is true both in terms of its relative frequency and its combinatory possibilities since in our analysis, it was common to all three semasiological clusters identified. However, is this feature typical of the semasiological structure of *annoy* or is it typical of some broader category, be that conceptual or functional? This problem would persist even if the data, in the form of personal diaries, were not biased towards first-person usage. In order to unequivocally establish the importance of the feature for the semasiological structure of the lexeme under investigation, it would have to be established that, relative to the correlations identified in the analysis, Patient – "1st Person" is significantly associated with this lexeme. Even if this were possible, would we need to compare the feature–lexeme frequency to all other verbs or only other verbs denoting emotions or only other verbs denoting concepts similar to *annoy*? This is an important question, because Patient – "1st Person" may be a prototypical feature of emotion concepts, just as it could be typical of the concept of ANNOY–BOTHER; this raises onomasiological questions, both relative to hyperonymy and synonymy of the lexeme. Indeed, the typicality of this feature may well be purely an epiphenomenal result of the fact that communication more generally is associated with first-person Patients.

Both the operationalisation of frequency-based prototypicality of a conceptual structure and the question of "semasiological salience" are examples of research topics where theory and method still need to come together. These two topics pair up in such a way that methodological issues reveal theoretical problems while theoretical problems are, of course, operationalised for quantitative analysis. It seems that for these two questions, both the theory and the method have some way to go before an adequate means for prototype-based description and explanation of semasiological structure has been achieved.

By way of conclusion, it is safe to say that Multifactorial Usage-Feature Analysis/the Profile-Based Approach can be applied to the study of polysemy. It does produce quantified falsifiable results and, importantly, its bottom-up approach to semasiological structure does permit a description in non-reified non-discrete terms. Given a frequency-based operationalisation of prototypicality, it is capable of distinguishing prototype effects in structure. However, properly applying fuzzy set theory to the analysis and properly interpreting these results in terms of prototype set theory requires further research.

Acknowledgement: The author wishes to sincerely thank two anonymous reviewers, the general editor Hubert Cuyckens, as well as the special issue editors for their observations, criticisms, and help. Evidently, all shortcomings remain my own.

References

Artstein, Ron & Massimo Poesio. 2007. Inter-coder agreement for computational linguistics. *Computational Linguistics* 34. 555–596.

Q5 Berez, Andrea & Stefan Th. Gries. 2009. In defence of corpus-based methods: A behavioral profile analysis of polysemous *get* in English. In Steven Moran, Darren Tanner & Michael Scanlon (eds.), *University of Washington Working Papers in Linguistics*, Vol. 27, 57–166.

Brugman, Claudia. 1983. *The story of over: Polysemy, semantics, and the structure of the lexicon*. Trier: LAUT.

Claude, Julien. 2008. *Morphometrics with R*. New York: Springer.

Coleman, Linda & Paul Kay. 1981. Prototype semantics: The English word *lie*. *Language* 57. 26–44.

Cuyckens, Hubert. 1995. Family resemblance in the Dutch spatial prepositions *door* and *langs*. *Cognitive Linguistics* 6. 183–207.

Deshors, Sandra. 2014. Constructing meaning in L2 discourse: The case of modal verbs and sequential dependencies. In Dylan Glynn & Mette Sjölin (eds.), *Subjectivity and epistemicity: Corpus, discourse, and literary approaches to stance*, 329–348. Lund: Lund University Press.

Q6 Deshors, Sandra. Forthcoming. *Multidimensional perspectives on interlanguage: Exploring* may *and* can *across learner corpor*a. Louvain: Presses Universitaires de Louvain.

Deshors, Sandra & Stefan Th. Gries. 2014. A case for the multifactorial assessment of learner language: The uses of may and can in French-English interlanguage. In Dylan Glynn & Justyna Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 179–204. Amsterdam: John Benjamins.

Dirven, René, Louis Goossens, Yvan Putseys & Emma Vorlat. 1982. *The scene of linguistic action and its perspectivization by SPEAK, TALK, SAY, and TELL*. Amsterdam: John Benjamins.

Divjak, Dagmar. 2006. Ways of intending: A corpus-based Cognitive Linguistic approach to near-synonyms in Russian. In Stefen Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics: Corpus-based approaches to syntax and lexis*, 19–56. Berlin: Mouton de Gruyter.

Divjak, Dagmar. 2010. *Structuring the lexicon: A clustered model for near-synonymy*. Berlin: De Gruyter Mouton.

Divjak, Dagmar & Nick Fieller. 2014. Cluster analysis: Finding structure in linguistic data. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative Cognitive Semantics: Corpus-driven approaches*, 405–442. Berlin: De Gruyter Mouton.

Divjak, Dagmar & Stefan Th. Gries. 2006. Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2. 23–60.

Fabiszak, Małgorzata, Anna Hebda, Iwona Kokorniak & Karolina Krawczak. 2014. The semasiological structure of Polish *myśleć* 'to think': A study in verb-prefix semantics. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative Cognitive Semantics: Corpus-driven approaches*, 223–252. Berlin: De Gruyter Mouton.

Fillmore, Charles. 1975. An alternative to checklist theories of meaning. *Proceedings of the Berkeley Linguistics Society* 1. 123–131.

Fillmore, Charles. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6. 222–254.

Geeraerts, Dirk. 1986. On necessary and sufficient conditions. *Journal of Semantics* 5. 275–291.

Geeraerts, Dirk. 1989. Prospects and problems of prototype theory. *Linguistics* 27. 587–612.

Geeraerts, Dirk. 1990. The lexicographical treatment of prototypical polysemy. In S. Tsohatzidis (ed.). *Meanings and prototypes: Studies in linguistic categorization*, 195–210. London: Routledge.

Geeraerts, Dirk. 1993a. Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics* 4. 223–272.

Geeraerts, Dirk. 1993b. Generalised onomasiological salience. *Belgian Journal of Linguistics* 8. 43–56.

Geeraerts, Dirk. 1995. Representational formats in Cognitive Semantics. *Folia Linguistica* 39. 21–41.

Geeraerts, Dirk. 2006. Methodology in Cognitive Linguistics. In Gitte Kristiansen, Michel Achard, René Dirven & Francisco J. Ruiz de Mendoza Ibañez (eds.), *Cognitive Linguistics: Current applications and future perspectives*, 21–50. Berlin: Mouton de Gruyter.

Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema. 1994. *Structure of lexical variation: Meaning, naming and context*. Berlin: Mouton de Gruyter.

Geeraerts, Dirk, Stefan Grondelaers & Dirk Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: Een onderzoek naar kleding- en voetbaltermen* [Convergence and divergence in Dutch vocabulary: A study of clothing and football terms]. Amsterdam: Meertens Instituut.

Glynn, Dylan. 2008. Lexical fields, grammatical constructions and synonymy: A study in usage-based Cognitive Semantics. In Hans-Jörg Schmid & Sandra Handl (eds.), *Cognitive foundations of linguistic usage-patterns*: *Empirical studies*, 89–118. Berlin: Mouton de Gruyter.

Glynn, Dylan. 2009. Polysemy, syntax, and variation: A usage-based method for Cognitive Semantics. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New directions in Cognitive Linguistics*, 77–106. Amsterdam & Philadelphia: John Benjamins.

Glynn, Dylan. 2010a. Testing the hypothesis: Objectivity and verification in usage-based Cognitive Semantics. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative Cognitive Semantics: Corpus-driven approaches*, 239–270. Berlin: De Gruyter Mouton.

Glynn, Dylan. 2010b. Corpus-driven Cognitive Semantics: Introduction to the field. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative Cognitive Semantics: Corpus-driven approaches*, 1–42. Berlin: De Gruyter Mouton.

Glynn, Dylan. 2014a. Polysemy and synonymy: Corpus method and cognitive theory. In Dylan Glynn & Justyna 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*: Corpus methods and Socio-Cognitive Semantics. In Dylan Glynn & Justyna Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 117–144. Amsterdam: John Benjamins.

Glynn, Dylan. 2014c. Techniques and tools: Corpus methods and statistics for semantics. In Dylan Glynn & Justyna Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 307–342. Amsterdam: John Benjamins.

Glynn, Dylan. 2014d. Correspondence Analysis: An exploratory technique for identifying usage patterns. In Dylan Glynn & Justyna Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 443–486. Amsterdam: John Benjamins.

Glynn, Dylan. 2014e. Conceptualisation of HOME in popular Anglo-American texts: A multifactorial diachronic analysis. In Javier Díaz-Vera (ed.), *Metaphor and metonymy across time and cultures*, 265–294. Amsterdam: John Benjamins.

Glynn, Dylan. 2014f. The social nature of ANGER: Multivariate corpus evidence for context effects upon conceptual structure. In Iva Novakova, Peter Blumenthal & Dirk Siepmann (eds.), *Emotions in discourse*, 69–82. Frankfurt am Main: Peter Lang.

Glynn, Dylan. 2015. Semasiology and onomasiology: Empirical questions between meaning, naming and context. In Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk Speelman & Hubert Cuyckens (eds.), *Change of paradigms – New Paradoxes: Recontextualizing Language and Linguistics*, 47–79. Berlin: De Gruyter Mouton.

Q7 Glynn, Dylan. Forthcoming. Cognitive Socio-Semantics: A quantitative study of dialect effects on the polysemy of *annoy*. *Review of Cognitive Linguistics*.

Glynn, Dylan & Kerstin Fischer (eds.). 2010. *Quantitative methods in Cognitive Semantics: Corpus-driven approaches*. Berlin: De Gruyter Mouton.

Glynn, Dylan & Karolina Krawczak. 2014. Operationalisation of non-observable usage-features: An exploratory study in English and Polish. Paper presented at *the International Conference on Evidentiality and Modality in European Languages*, Madrid, 6–8 October.

Glynn, Dylan & Justyna Robinson (eds.). 2014. *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*. Amsterdam: John Benjamins.

Goguen, Joseph. 1967. L-fuzzy sets. *Journal of mathematical analysis and applications* 18. 145–174.

Goguen, Joseph. 1969. The logic of inexact concepts. *Svnthese* 19. 325–373.

Greenacre, Michael 2007. *Correspondence analysis in practice*, 2nd edn. Boca Raton: Chapman & Hall.

Gries, Stefan Th. 1999. Particle movement: A cognitive and functional approach. *Cognitive Linguistics* 10. 105–145.

Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. Continuum: London.

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

Gries, Stefan Th. 2010. Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon* 5. 323–346.

Gries, Stefan Th. & Dagmar Divjak. 2009. Behavioral profiles: A corpus-based approach towards cognitive semantic analysis. In Vyvyan Evans & Stephanie Pourcel (eds.), *New directions in Cognitive Linguistics*, 57–75. Amsterdam: John Benjamins.

Gries, Stefan Th. & Naoki Otani. 2010. Behavioral profiles: a corpus-based perspective on synonymy and antonymy. *ICAME Journal* 34. 121–150.

Gries, Stefan Th. & Anatol Stefanowitsch (eds.). 2006. *Corpora in Cognitive Linguistics: Corpus-based approaches to syntax and lexis*. Berlin & New York: Mouton de Gruyter.

Heider, Eleanor [Rosch]. 1971. 'Focal' color areas and the development of color names. *Developmental Psychology* 4. 447–455.

Heider, Eleanor [Rosch]. 1972. Universals in color naming and memory. *Journal of Experimental Psychology* 93: 10–20.

Herskovits, Annette. 1986. *Language and spatial cognition: An interdisciplinary study of the prepositions in English*. Cambridge: Cambridge University Press.

Heylen, Kris, Thomas Wielfaert, Dirk Speelman & Dirk Geeraerts. 2015. Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua* 157. 153–172.

Hopper, Paul. 1987. Emergent grammar. *Berkeley Linguistics Society* 13. 139–157.

Horvath, Steve. 2011. *Weighted network analysis: Applications in genomics and systems biology*. New York: Springer.

Janda, Laura. 1990. Radial network of a grammatical category – its genesis and dynamic structure. *Cognitive Linguistics* 1. 269–288.

Klavan, Jane. 2014. A multifactorial corpus analysis of grammatical synonymy: The Estonian adessive and adposition peal 'on'. In Dylan Glynn & Justyna Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 253–278. Amsterdam: John Benjamins.

Krawczak, Karolina. 2014a. Corpus evidence for the cross-cultural structure of social emotions: Shame, embarrassment, and guilt in English and Polish. *Poznań Studies in Contemporary Linguistics* 50. 441–475.

Krawczak, Karolina. 2014b. Epistemic stance predicates in English: A quantitative corpus-driven study of subjectivity. In Dylan Glynn & Mette Sjölin (eds.), *Subjectivity and epistemicity: Corpus, discourse, and literary approaches to stance*, 355–386. Lund: Lund University Press.

Krawczak, Karolina & Dylan Glynn. 2015. Operationalising mirativity: A usage-based quantitative study on constructional construal in English. *Review of Cognitive Linguistics* 13(2). 253–282.

Krawczak, Karolina & Iwona Kokorniak. 2012. A corpus-driven quantitative approach to the construal of Polish *think*. *Poznań Studies in Contemporary Linguistics* 48. 439–472.

Q8 Labov, William. 1973. The boundaries of words and their meanings. In Joshua Fishman (ed.), *New ways of analyzing variation in English*, 340–373. Washington: Georgetown University Press.

Lakoff, George. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic* 2. 458–508.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Langacker, Ronald. 1987. *Foundations of Cognitive Grammar*. Vol. 1: *Theoretical prerequisites*. Stanford: Stanford University Press.

Priness, Ido, Oded Maimon & Irad Ben-Gal. 2007. Evaluation of gene-expression clustering via mutual information distance measure. *BMC Bioinformatics* 8. 111

Rastier, François. 1987. *Sémantique interprétative*. Paris, Presses Universitaires de France.

Rosch, Eleanor [nee Heider]. 1973. Natural categories. *Cognitive Psychology* 4. 328–350.

Rosch, Eleanor [nee Heider]. 1975. Cognitive reference points. *Cognitive Psychology* 7. 532–547.

Rudzka-Ostyn, Brygida. 1985. Metaphoric processes in word formation. In Wolf Paprotté & René Dirven (eds.), *Ubiquity of metaphor: Metaphor in language and thought*, 209–241. Amsterdam: John Benjamins.

Rudzka-Ostyn, Brygida. 1988. Semantic extensions into the domain of verbal communication. In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, 507–553. Amsterdam: John Benjamins.

Rudzka-Ostyn, Brygida. 1989. Prototypes, schemas, and cross-category correspondences: The case of *ask*. *Linguistics* 27. 613–661.

Rudzka-Ostyn, Brygida. 1995. Metaphor, schema, invariance: The case of verbs of answering. In Louis Goossens, Paul Pauwels, Brygida Rudzka-Ostyn, Anne-Marie Simon-Vandenbergen & Johan Vanparys (eds.), *By word of mouth: Metaphor, metonymy, and linguistic action from a cognitive perspective*, 205–244. Amsterdam: John Benjamins.

Sandra, Dominiek & Sandra Rice. 1995. Network analyses of prepositional meaning: Mirroring whose mind – the linguist's or the language user's? *Cognitive Linguistics* 6. 89–130.

Schütze, Henrich. 1998. Automatic word sense discrimination. *Computational Linguistics* 24. 97–123.

Speelman, Dirk & Dylan Glynn. 2005. *LiveJournal corpus of American and British English*. Leuven: University of Leuven, Department of Linguistics.

Talmy, Leonard. 1985. Force dynamics in language and cognition. *Cognitive Science* 12. 49–100.

Taylor, John. 1989. *Linguistic categorization: Prototypes in linguistic theory*. Oxford: Clarendon Press.

Turney, Peter & Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37. 141–188.

Tyler, Andrea & Vyvyan Evans. 2001. Reconsidering prepositional polysemy networks: The case of *over*. *Language* 77. 724–765.

UNESCO. 2013. Statistical guide for partitioning around medoids. http://tinyurl.com/yh5qvqj (accessed 25 April 2016)

Vandeloise, Claude. 1986. *L'espace en français*. Paris: Seuil.

Victorri, Bernard. 1997. La polysémie: Un artéfact de la linguistique? *Revue de sémantique et pragmatique* 2. 41–62.

Wierzbicka, Anna. 1990. Prototypes save: On the uses and abuses of the notion "prototype" in linguistics and related fields. In Savas L. Tsohatzidis (ed.), *Meanings and prototypes*, 347–367. London: Routledge.

Zadeh, Lofti 1965. Fuzzy sets. *Information and Control* 8. 338–353.

Zadeh, Lofti 1968. Probability measures of fuzzy events. *Journal of Mathematical Analysis and Applications* 23. 421–427.