The many uses of *run*

Corpus methods and socio-cognitive semantics

Dylan Glynn
Lund University

1. Corpus data, usage features, and prototype structure

Gries' (2006) study 'Corpus-based methods and cognitive semantics: The many senses of *to run'* counts amongst the most influential contributions to the study of polysemy in Cognitive Linguistics. It is important not because its methodology is original, nor because it is complete or extensive, nor even because of the theoretical claims it makes, but because it simply and overtly shows how the corpus-driven usage-feature method can be applied to the study of polysemy and prototype structuring. Its contribution is the combination of theory, model, and method - three pieces of a puzzle that establish the foundations of a theoretically and empirically coherent approach to semasiological structuring. The current study, in contrast, is of a very simple nature. It will not challenge the theory, the method, or the results of Gries (2006), it will simply suggest an addendum; an afterword that demonstrates the simple, but important point, that semantic structure is influenced by social variation.

The study begins with a brief theoretical discussion, considering the strengths and weaknesses of radial network analysis (section 1.1). It explains the goals of corpus-driven usage-feature analysis and argues the importance of context in semantic analysis (1.2). Section two describes the case study and repeats the analysis presented in Gries (2006). It then moves to an empirical demonstration of the need to include the social dimension in semantic research.

1.1 Usage-feature analysis and radial network analysis

Lakoff (1990) presented the commitment to empiricism and inductive research as the gold standard of Cognitive Semantics. Although prototype category theory and the radial network analysis that employs it (Lakoff 1987) represented necessary and substantial steps toward an empirical Cognitive Semantics, it did not quite attain that goal.[1] Shortcomings not withstanding, the research of the era demonstrated that:

(i) linguistic semantics, as opposed to pragmatics, cannot adequately account for meaning structure in language - instead, it demostrated the need for 'encyclopaedic semantics'

(ii) necessary and sufficient conditions cannot adequately determine socio-conceptual categories - instead, it demonstrated the need for 'prototype effects'

Radial network analysis (Lakoff 1987), as well as Frame Semantics (Fillmore 1985), was an analytical model designed to offer a rigorous means for describing meaning structure assuming both prototype effects and encyclopaedic semantics. How successful was this analytical model? Methodologically, radial network analysis continued the Structuralist and Generativist tradition:

(i) empirically, network analysis employed introspection to determine language structure;

(ii) analytically, network analysis assumed this structure to take the form of discrete senses.

Seen from this perspective, Cognitive Semantics made the first important steps towards empiricism, but it did not reach its goal. Theoretically and empirically, the shortcomings are well established (Geeraerts 1993; Sandra & Rice 1995). The usage-feature, or profile-based, methodology developed by Geeraerts *et al.* (1994) and Gries (2003) *inter alia*, is, however, a contribution to that endeavour.

There is no need to re-cover well-trodden ground. Let us assume the theoretical models of encyclopaedic semantics (Fillmore 1985; Lakoff 1987) and prototype categorisation (Rosch 1975; Lakoff 1987). The first replaces truth-conditional semantics with world knowledge. The second replaces necessary and sufficient conditions with prototype structure. These two proposals accepted, we can turn to the analytical assumption of discrete senses and the methodological technique of introspection.

Firstly, the assumption of discrete senses is something that is intuitively attractive. Indeed, it is as obvious as the world is flat - words have meanings and we choose between those meanings in communication. These meanings are, therefore, reified or objectified as discrete units. This naïve operationalisation certainly aids in language learning, dictionary writing and, typically, only comes amiss in inter-personal disputes. However, the evidence for discrete lexical senses is as naïvely sound as the horizon is horizontal. Since Fuzzy Set Theory and Prototype Set theory both dispose of discrete conceptual categories (necessary and sufficient conditions), why do we continue to assume that the conceptual sub-categories (lexical senses) are discrete?

Of course we do not assume discrete senses in theory (we return to this below), but radial network analysis does in practice. Lakoff's (1987) study of *over* identifies a list of usage-features in terms of minimal perceptual distinctions expressed as schemata. Rather than seeing meaning construction as the relative correlation of schema features, Lakoff continues to hunt for 'senses' as reified configurations of those schema features. Even if for perceptual categories like OVER discrete senses are ultimately possible, this is unlikely to be the case for lexical semantics as a whole, let alone grammatical semantics. Let us assume, therefore, that discrete senses are a useful notion in discussing semasiological structure in applied linguistics, but let us not assume that reified discrete senses actually exist.

Secondly, radial network analysis employed introspective methodology. Although introspection has an essential and inarguable role in language research, both for proposing hypotheses and performing analyses, with no truth-conditional tests to help determine conceptual structure, it will only ever be one part of an empirical science. Tyler & Evans (2001) have attempted to develop a 'principled approach' to identifying semasiological structure. This goes a long way towards minimising the risk of *ad hoc* categorisation using introspection. It does not, however, offer the possibility of result falsification. It is this second point that is essential. According to their own models of language, Generativists and Structuralists both had means for falsification using introspection. However, employing a usage-based theory of language means that introspection is severely limited as an analytical method.

Accusing Cognitive Semantics of relying solely on introspection and of assuming the existence of discrete senses is, perhaps, unfair. In the study of polysemy alone, there exist both corpus and experimental traditions. Indeed, the corpus-driven usage-feature approach propounded by Gries (2006) goes back to the very origins of Cognitive Semantics (Dirven *et al* 1982), just as elicitation-driven usage-feature analysis (Lehrer 1982) also finds itself at the origins of the theoretical paradigm. Moreover, as early as Geeraerts (1993), Geeraerts *et al*. (1994), and Lehrer & Lehrer (1994) in both theoretical and empirical terms, the argument for a non-reified approach to lexical senses was put forward. Therefore, even if it did not represent the main drive of research, Cognitive Semantics has slowly been moving towards empirical methods and some in the field have long held that meanings cannot be understood as reified objects.

In this vein, Gries (2006) is but one empirical Cognitive Semantic study in a long history. Its step was to begin to develop a descriptive model based on multifactorial statistics for dealing with the non-reified structuring that corpus-driven usage-feature analysis identifies. This step, interpreting multifactorial results in terms of prototype effects, is important. Empirical methods and non-reified senses may be theoretically sound, but with no way of modelling the results or coherently representing the structuring of language, the application of the method will struggle to gain ground. Therefore, a corpus-driven usage-feature methodology should:

a. produce results that can be empirically falsified;
b. posit socio-cognitively realistic semantic structures;
c. capture relative frequency-based prototype effects.

1.2 Prototype structure and methodological reasons for context

Gries' (2006: 75-77) study introduces the notion of prototype structure in corpus-driven polysemy research. To appreciate the ultimate argument of our discussion, we must return to the question of prototype structuring and develop certain points. Gries (2006: 75) offers a "non-exhaustive" list of different operationalisations of the notion of polysemy: intuition determined judgements of similarity and 'goodness'; elicitation ease; diachronic evidence; centrality/predominance in a radial network, and so forth. We have no issue with this nor with the fact that the proposed prototype is justified by comparing the frequency-based results with several of these other operationalisations (Gries 2006: 76).

For the semasiological variation of a term such as *run*, we can suppose there would be little debate that 'fast pedestrian motion' is the prototypical 'meaning'. From synchronic frequency of use, diachronic evidence of earliest uses, and intuition-based conceptual salience, to widely accepted theories of embodiment and primacy of perception, all evidence points unanimously to 'fast pedestrian motion' as the 'central', and probably most 'typical', meaning. However, we must consider two issues.

The first issue is the operationalisation of polysemy. Prototype structure is an analytical model; it is not an object of study. It can be used to explain different structures in language, depending on how it is operationalised. Geeraerts (1987: 288) notes that there are two basic operationalisations of polysemy. He terms these the analytic and introspective criteria of prototypicality. Although his debate was with Structuralism and truth-conditional semantics, we can rephrase this, *mutatis mutandis*, as frequency-based versus salience-based prototypicality. There are many different approaches to prototype structure, but it is likely that they will all be based on either one of these two operationalisations. Perceptual – conceptual 'focus' versus relative frequency 'commonness' are two fundamentally different ways of approaching socio-cultural typicality and prototype effects. As Geeraerts (1987) shows, theoretically, there is no reason to assume that prototype models using one or the other operationalisations would offer the same results. This is not to say they will not. For an example as conceptually simple as *run*, it is likely they will and this is why in Gries' (2006: 76) comparison of different methods, each method indicates the same prototype structure. However, if we are developing a methodology for identifying semantic structure, it is important we do not make the assumption that these different methods should necessarily offer convergent results.

Schmid (2000) and Gries (2003) have both made claims about the relationship between frequency and conceptual structure. These claims have yet to be confirmed empirically and the authors appear to have distanced themselves from their earlier position (Schmid 2010, Gries p.c). Although frequency of occurrence surely has an important role in determining conceptual structuring, conceptual and perceptual salience are also likely to have an impact. It is, therefore, unlikely that there is a one-to-one index where more frequent equates more central. Issues that ensue from trying to compare such results from these two different methods are beginning to take centre stage. Arppe & Järvikivi (2007), Arppe *et al.* (2009), Tribushinina (2009), and Gilquin (2010) are amongst recent examples of research that is seeking to understand how these two fundamentally different notions of prototypicality interact. Eventually, we may understand how their interaction impacts upon language structure and learning, but for the moment, this has not been determined.

The second issue is related to the first in that it comes from treating prototypes as an object of study rather than an analytical model. The circularity of that is not unlike the circularity of previous approaches to language that seek to use a model to prove the accuracy of the model. One cannot use Prototype Set Theory to prove prototype structuring of language any more than one can use Government and Binding Theory to prove the syntactic structuring of language. Gries is perfectly aware of this and one of the reasons why his study is important is precisely because he proposes a method by which we can test the hypotheses generated employing the prototype model of semantic structure, represented in the radial network research. In brief, he does not use the model to produce the results, he compares the results to what the model would predict, fulfilling Lakoff's (1990) goal of generalisation (induction) based on empirical research.

However, it must be remembered that the two senses that Gries identifies as the most frequent, and therefore (proto)typical, were predicted by a set of usage-features (Gries 2006: 85). So it is not the senses, but the 'configurations of features', to use the terminology of Geeraerts *et al.* (1994) or the 'behaviour profiles of ID tags' to use Gries' terminology, that are the prototype structures. This is why we should not speak of the many 'senses' of *run* but of the many 'uses', where use is operationalised as a configuration of usage-features (or ID-tag profiles). This does not contradict Gries' results, but further emphasises their theoretical and methodological implications.

What are these implications? That Gries is able to predict the occurrence of the two most common 'senses' is interesting, but we need to invert the logic to see its importance. This means that we have an operational definition for these senses. The senses are predicted using a configuration of usage-features. Therefore, we can 'define' the senses as configurations of usage-features (or ID-tag profiles).

Yet the implications go further than providing an operational definition. These usage-features are, in fact, contexts. Given the context of tense (past), transitivity (transitive), complement syntax (*to* + infinitive), and agent type (Human), Gries is able to predict with 100% accuracy the 'fast pedestrian motion' sense of *run* (in the dataset used in the study). In itself, this does not mean it is prototypical, merely that it has a clear behavioural profile, and that its usage pattern can be clearly identified. It is its frequency that Gries argued to be the reason for its (proto)typicality. In terms of frequency, the features of 'past tense', 'transitive', 'finite *to* infinitive syntax', and 'human agent', were also the most frequent. Therefore, we have an operationalisation of the definition of a sense of *run*, but also an operationalisation of its typicality, or frequency-based prototypicality.

This discussion has sought to flesh out some of the theoretical and methodological implications of Gries' study, but it has also introduced the notion of context. We can now move on to consider this notion.

Given the two operationalisations above (sense – as the frequency of the configuration of features and (proto)typicality – as the frequency of those configurations), we see why, methodologically, we need context. Geeraerts (2000) proposed what he terms onomasiological and semasiological salience. This point is crucial in frequency-based studies of semantic structure. If we assume an analytical criterion for polysemy study, described by Geeraerts (1987) and propounded by Lakoff (1990), as opposed to an abstracted *langue* based approach to conceptual categorisation, then frequency (typicality and possibly centrality) is likely to be entirely context dependent. Let us take Gries' two frequency-based typical meanings 'fast pedestrian motion' and 'manage'. In the context of children, 'fast pedestrian motion' will be extremely frequent and 'manage' extremely infrequent. By contrast, in the context of economic news press the 'fast pedestrian motion' is likely to be extremely infrequent, especially compared to 'manage'.

Are we trying to determine a typical meaning that is true for all language in all contexts? If this were possible (especially for a language as diverse as English), it is surely not possible with any corpus currently available or available in the foreseeable future. Even taking a single context distinction, spoken versus written, the largest and most 'balanced' corpus in existence is non-representative to an unimaginable degree. The quantity of written language compared to spoken is infinitesimally small. Note, however, that this ignores the fact that some written language has a much greater conceptual 'impact' upon the speech community than much spoken language. We must ignore this because we are working in terms of frequency and still do not have a clear understanding of how the two frequency-based operationalisations (sense - frequency of features in configuration, and typicality - frequency of configurations of features) inform salience and / or conceptually based operationalisations of sense and typicality.

In brief, the study of frequency-based prototype effects must be relative to context. We, therefore, must posit (proto)typicality structures, not for an entire language but for a language context. Saussure (1997[1916]: 124[184]) identified the *langue* as a point on a two-dimensional Cartesian plane. It is often forgotten that a point on that plane not only represents variation along the *y*-axis, but that synchronic plane also varies relative to register, dialect, age, gender and enumerable different shades of context. Therefore, let us not employ usage-based methods to describe the hypothetical and reductionist *langue* of Structuralism or *competence* of Generativism. Our object of study is synchronically varied – our models of conceptual structure must also be so.

2. Case study: *run* in America and Britain in diaries and conversation

2.1 Two corpus-based studies on *run*

Our current study imitates Gries (2006) in the set of usage-features (ID tags) analysed as closely as possible. The aim is not to test the results or to improve upon them through more advanced statistical analysis or a larger, more diverse sample. The aim is merely to show that even for a lexeme as culturally 'simple' and as sociolinguistically 'neutral' as *run*, one must account for the social dimension of language in semantic analysis.[2] We begin with a a summary of Gries' (2006) study.

Gries' analysis is based on 815 occurrences of the lemma *to run*, extracted from the British component of the International Corpus of English and the Brown Corpus of American English. Approximately 400 occurrences were taken from each. These occurrences were manually analysed and categorised (using intuition) as belonging to one of 48 senses. These senses were taken from the Collins Cobuild E-Dictionary, the Merriam Webster's American online dictionary and the WordNet project. This categorisation in terms of dictionary senses is the first factor of the analysis. Although it is normally the goal of usage-feature analysis to determine different 'senses' through the identification of 'feature configurations' (Geeraerts *et al.* 1994) or 'behavioural profiles' (in Gries' terminology), being able to match such configurations against dictionary definitions is a useful heuristic. In Gries (2006), it is used to show how a frequency-based study can inform an understanding of prototype structure in polysemy.

The 815 occurrences in Gries' data set are analysed for a range of factors, or usage dimensions. These factors consist of the usage-features typical in this kind of methodology - formal and semantic features ranging from syntax and collocation, tense and aspect, to the semantics of the argument structure and participants. In this study, the formal factors include tense, aspect, voice, transitivity, mood, and clause type. The semantic factors include subject type, object type, and complement type. These type features are categories such as human, concrete countable object, concrete mass noun, machines, abstract entities, organisations, locations, quantities, events, processes, etc.

The dictionary senses found are exemplified and enumerated. The most frequent dictionary senses identified are that of ‘fast pedestrian motion’ (203 occurrences / 25%, exemplified p.63) and ‘manage’ (101 occurrences / 12% exemplified p. 71). The analysis and subsequent categorisation of the occurrences as dictionary senses is systematically explained by example. It is this systematic explanation that is used in the current study to repeat the analysis and categorise the occurrences as dictionary definitions.

The current study is based on 500 occurrences of *run*, 250 each of British and American English, subdivided again into 125 examples each of conversation and online personal diary. The sample was restricted to this relatively small number due to practical reasons – usage-feature analysis is laborious and resource consuming. The point of the study being to investigate the need to include sociolinguistic parameters in polysemy research, the improved descriptive accuracy afforded by increasing this number would not substantially improve the ability to demonstrate the point. Also, the methods under investigation must be shown to produce coherent results with small numbers, since, for the same practical reasons, the usage-feature (or profile-based) method, tends to deal with small samples. The British and American diary examples were taken from the LiveJournal corpus, developed by D. Speelman, at the University of Leuven, and the conversation examples were taken from the British National Corpus and the American National Corpus. The usage-feature analysis is replicated using the same dictionary senses employed by Gries and the same range of formal and semantic usage-features.

An aside should be made here. Despite the fact that Gries more than adequately demonstrates the principle of the method, descriptively, the study is preliminary (Gries 2006: 81). The obvious question of why one would focus on dictionary senses (instead of soley usage-features, or ID-tags, to use Gries’ terminology) can be answered by the fact that the study’s aim is to show how prototype structure can be handled with the method. Nevertheless, in terms of descriptive adequacy, this option is far from ideal. Moreover, as the author stresses himself, the size of the sample is too small to properly apply multivariate statistical analysis. It is not that the sample is small in itself, but the type-token (or perhaps ‘sense-token’) ratio is not acceptable for multifactorial analysis. Gries repeatedly stresses this point, but it should be added that this problem is compounded by the fact that the study is not restricted to *run*, but includes all the verb particle constructions based on *run*. Arguably, this makes the study partially one of near-synonymy instead of polysemy. Many of the senses identified are determined formally by the combination of the verb and the particle. Verb particle constructions in Germanic, just like the prefixed verb constructions in Slavic (see Fabiszak *et al*., this volume), challenge the distinction between synonymy and polysemy. In any case, many of the senses in question are both formally and semantically distinct. A true test of the usage-feature method for the study of semasiological variation is when that variation is not linked to any overt, or obvious, formal distinction. By excluding the verb particle construction, Gries' study would have included less semantic variation but also less formal variation for 'automatically' determining it. This does not detract from the goal or the results of Gries' study, but future work should take such questions into account. Note that the current study also uses dictionary senses as one of its analytical factors and includes the particle constructions. This is done to permit a comparison with Gries.

Table 1, below, lists the most common senses in the current study, compared with the figures from Gries (2006). The list of senses applied in this study was determined by the senses submitted to the hierarchical cluster analysis in Gries (2006: 82). For some of these senses, the number of occurrences (supplied in the preceding section, Gries 2006: 63-73) are not known. Although the reasoning behind the categorisation of the examples as dictionary definitions is reasonably clear, taxonomical issues of hyperonymy in the discussion occasionally mean the number of occurrences for a given sense is not stated. This is the case for 'function' vs. 'execute' and 'manage' and for 'free motion' versus 'motion' and 'fast motion'.

The application of Gries' dictionary senses to our data was reasonably straightforward, using the examples and explanations included in the study. There were, of course, some classification issues. For example, what constitutes 'fast' in 'fast motion'? The large difference in the number of occurrences on this point suggests that there may have been a difference in coding for this sense. Nevertheless, assuming there is bound to be some analytical variation, the results are reasonably comparable. This is especially true seeing the small size of the samples and differences between corpora. The principal differences are 'become used up' and 'escape', which are more frequent in this study, and 'manage', which is substantially more frequent in Gries' study. For this final difference, even if we allow for some confusion over the semantically similar categories of 'execute', 'in charge of', 'function' and 'manage', the difference is marked. We can suppose that such differences are a result of register. Indeed, this is precisely the problem with frequency-based studies in typicality. Thematic variation, or variation in 'topic of discourse', can have a substantial effect, even upon coarse-grain analysis of semasiological structure. Table 1, below, lists the five most frequent literal, or 'motion' based senses, as well as the five most common figurative, or non-motion based senses, in both studies.

There is no need to examine such differences and similarities further. Both samples are small, with a high type-token ratio, which means statistical significance would tell us little. Remembering that the ultimate point is to show that frequency-based prototype structures are context dependent, it is sufficient to show that the overall study is comparable to that of Gries'.

Table 1. Most frequent dictionary senses

Dictionary sense	Current study	Gries (2006)
'fast pedestrian motion'	160 (32%)	203 (25%)
'escape'	57 (11.5%)	32 (4%)
'motion'	23 (4.5%)	24 (3%)
'fast motion'	17 (3.5%)	4 (0.5%)
'free motion'	17 (3.5%)	-
'execute'	18 (3.5%)	28 (3.5%)
'in charge of'	16 (3%)	24 (3%)
'manage'	25 (5%)	101 (12%)
'function'	17 (3.5%)	-
'become used up'	26 (5%)	14 (2%)

Below is an exemplified list of the most common senses. The examples are all extracted from the LiveJournal sample under investigation. For further exemplification and discussion, see Gries (2006: 63-73).

(1) 'fast pedestrian motion'
I want to like run into a bathroom at school and cry my eyes out whenever i see him.

(2) 'escape'
does anyone have about $400 laying around, i think i want to run away to Las Vegas for a few days, lol.

(3) 'motion'
Action Cat is really starting to like the new kitty, who I call Buddy cause he has yet to recieve a formal name. They run around and play all the time now and it's really cute.

(4) ‘fast motion’
Hang on, till I get the brake on, or you'll run into the river.

(5) ‘free motion’
we've made three different trips ... the group of friends that i run around with.

(6) ‘execute’
you know like it's easier for you to go and run a program you know through the disk.

(7) ‘in charge of’
there's er it was for the er cat scanner and it was run by the Co-Op it was, it was just oh I saw that sign outside.

(8) ‘manage’
I am now the new landlord of the rose and crown pub which mama used to run.

(9) ‘function’
they said that uh cars would cost two dollars and they would run forever.

(10) ‘become used up’
Well, it doesn't do so bad. It's usually cigs we run out of not petrol.

2.2 Semasiological clustering without social dimensions

Gries (2006: 81-82) submits all the senses (minus ‘idiomatic’ ones) to an agglomerative hierarchical cluster analysis (see Divjak, this volume, for an explanation of the technique). The senses are clustered using the full range of features (Gries 2006: fn 19, p. 94). The results of cluster analysis are reasonably coherent, especially given the number of senses versus the number of examples and the number of usage-features. There is some degree of intuitively sound clustering, which could be re-interpreted as prototype structuring. Nevertheless, there is also a large amount of clustering that does not appear semantically motivated. Gries (2006: 81, 83) accepts this and suggests that the data sparseness is, at least partially, to blame.

Replicating the procedure gives similar results - a reasonable degree of intuitively sound clustering but also a reasonable amount of ‘noise’ in the dendrogram where clusters make little or no sense. For sake of brevity, we will not present the dendrogram, but instead present the cluster results obtained by simplifying the data and limiting the usage-features used to cluster them. In order to obtain a more coherent clustering of senses, rare senses were omitted. Also, the two most frequent senses, 'fast pedestrian motion' and 'escape' were omitted. These two senses were found to systematically dominate the clustering, rendering the relations between the other senses difficult to discern. We can suppose that these two senses were so distinct in usage that the clustering could not model their relationship and more subtle relations of the other senses simultaneously. This effect was found, regardless of the distance measure used.

Gries (2006: fn 19, pp. 93-94) found that changing the distance matrix and/or the agglomerating method did not substantially alter the results. This was not the case with the current data set. Experimenting with different agglomeration methods greatly improved or worsened the interpretability of the dendrogram and, occasionally, the actual results of the cluster analysis. Likewise, different distance measures also produced different results. We return to this below.

As mentioned, in an effort to obtain a simpler, more coherent clustering of the senses, the range of usage features submitted to the analysis was reduced. Gries shows that, formally, transitivity is one of the most important factors in the study. This factor was therefore retained. The argument type semantic factors, subject semantics and object semantics, were also retained. The rest of the factors were not submitted to the cluster analysis. Figure 1, below, presents the results using the simplified range of senses and the three usage-feature factors. It is produced using the Euclidean distance measure (the simplest distance measure), and 'average' as the agglomeration method (a common agglomeration method).

We must interpret such plots with caution. Even having removed the rarely occurring senses, some of the remaining senses are still infrequent, for example - 'caused motion', 'motion into difficulty' and 'campaign'. The overall picture seems reasonably coherent. Examining the dendrogram, two broad sense clusters emerge, clustered by the right and the left branches. The left branch includes most of the abstract senses, with perhaps the exception of 'function', which is less abstract. Note, however, that the analysis has 'function' and 'diffuse' as quite distinct from this abstract cluster. It appears the analysis has trouble incorporating these senses.The intuitive adequacy of the model is left up to the reader, but it is worth pointing out that the literal motion senses are coherently grouped together as well as the control senses ('manage', 'in charge of', 'execute'). However, the place of 'become used up' with these two groupings of senses is not clear, nor is the relationship between the 'control' senses and the literal motion senses.

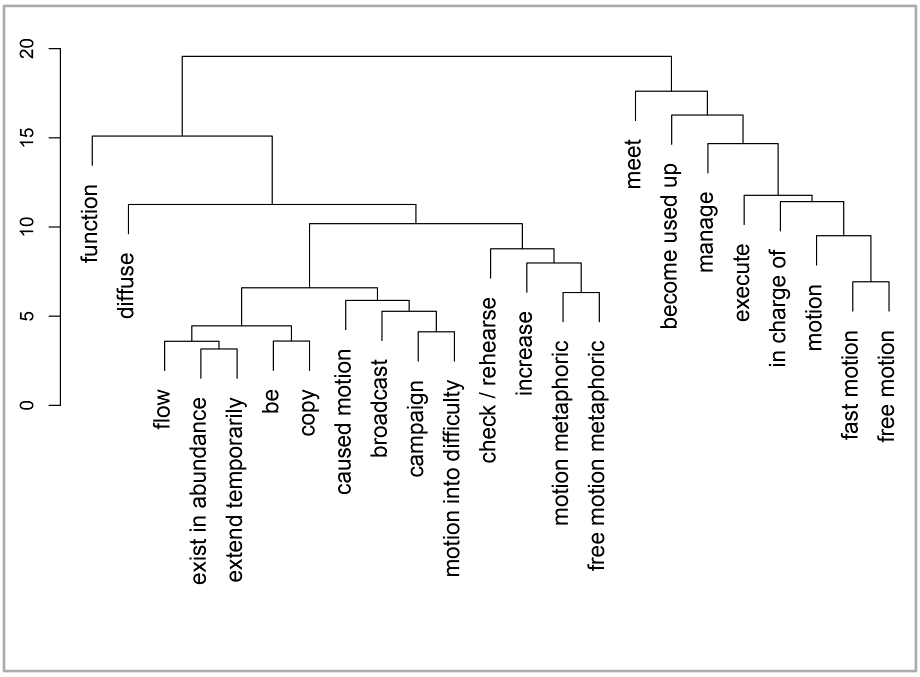

Figure 1. Hierarchical cluster analysis of dictionary senses
Distance matrix - Euclidean; agglomeration method - 'average'

There does exist an internal logic to the cluster of abstract senses. The metaphoric motion senses are grouped together, just as are the 'spread' senses of 'flow', 'exist in abundance', and 'extend temporarily'. The other groupings are not illogical, but apart from representing abstract or metaphoric meanings of *run*, they share little semantically.

As mentioned, variation in the results was found depending on the distance measure employed. This could, perhaps, be a sign of the instability of the analysis - attempting to cluster 23 senses based on a sample of 500 is far from ideal conditions in multivariate statistics. Figure 2 presents the results of the Canberra distance matrix. It is clustered with the Ward agglomeration method. The different agglomeration methods did not change the results for the Canberra matrix, only legibility. The Ward method gave the clearest dendrogram.

In Figure 2, again we see two main branches. At first, the results seem more coherent than those produced using the Euclidean distance measure. However, if we inspect the clustering more closely, intuitive semantic coherence is not wholly systematic. At the coarse grain level, we have lost the clear distinction between relatively concrete uses such as 'manage' and 'literal motion' versus 'metaphoric motion' as well as the extending and disseminating senses. For the four sub-clusters, there is a little more semantic coherence. The first sub-cluster of 'execute' and 'manage' is intuitively sound. The second is also coherent, save for the sense 'in charge of'. However, this is semantically related to 'execute' and 'manage' and, therefore, given the small sample, is more or less in the 'correct' branch. The next sub-cluster of 'difficulty' (run into difficulty), 'campaign' (run for election) and 'meet' (run into a friend) is semantically coherent, given a broad interpretation of 'campaign' that includes meeting people and difficulties. This is not as unlikely an interpretation as one might first suppose. Recall that the different semantic types of objects and subjects determine these sense clusters.

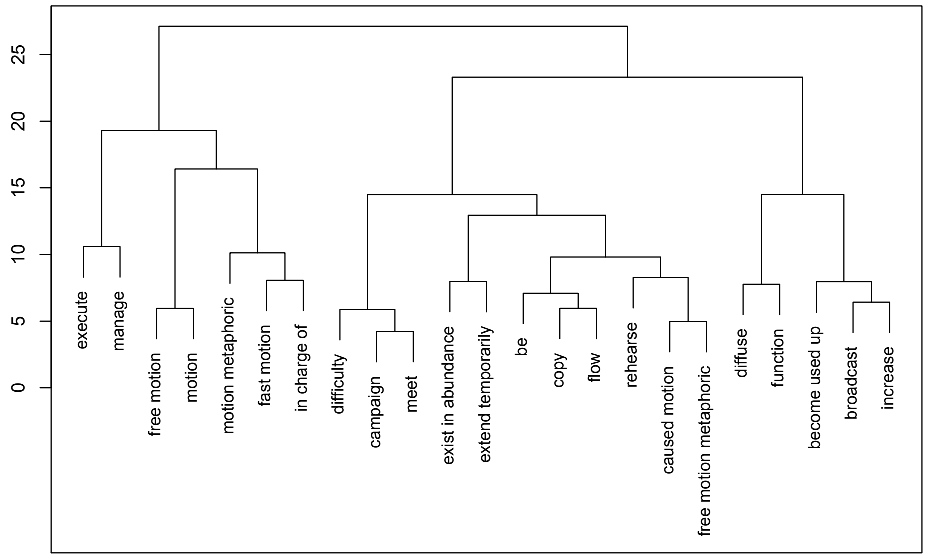

Figure 2. Hierarchical cluster analysis of dictionary senses
Distance matrix - Canberra; agglomeration method - 'Ward'

Moving to the right across the clusters, the next sub-cluster of 'exist in abundance' and 'extend temporarily' is intuitively coherent. However, the rest of the group appears semantically heterogeneous. The last cluster on the right, although distinct with a long branch stemming from the rest of the dendrogram, also lacks obvious semantic coherence. Although one is able interpret semantic structure here, it is not self-evident why 'diffuse' and 'function' or 'broadcast' and 'increase' should group together.

The point of both this small study and Gries' is merely to consider two methodological possibilities. In light of this, that the two distance matrices produced different clusterings raises important methodological questions. Standards and checks for appropriateness need to be developed before the use of cluster analysis can be relied upon to determine frequency-based semasiological structure.[3]

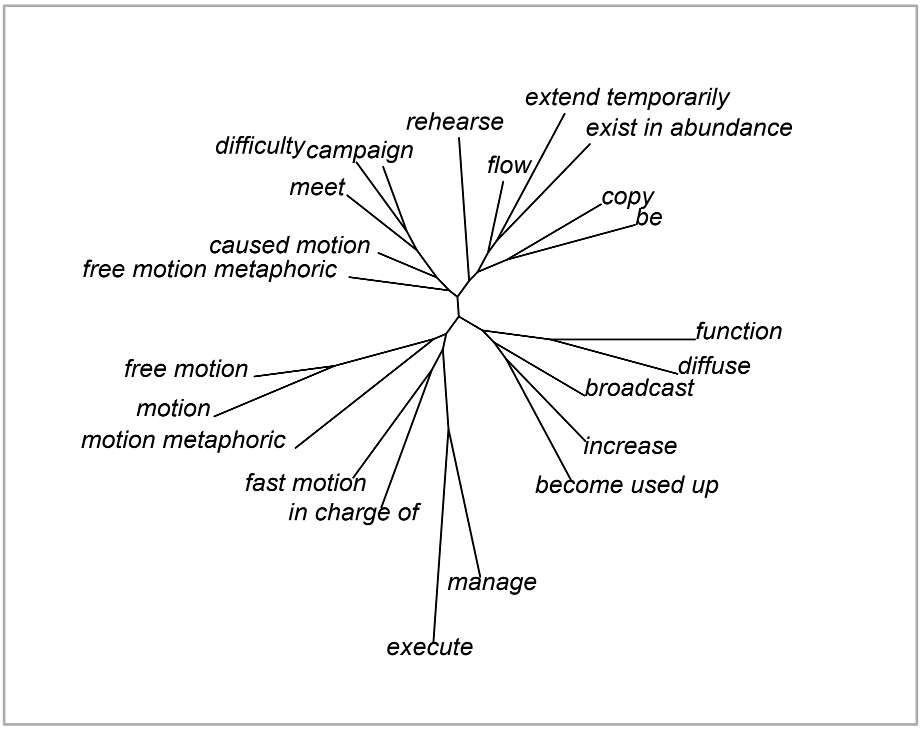

Figure 3. Unrooted cluster analysis of dictionary senses
Distance matrix - Canberra; method - 'phylogenetic'

To conclude this section, we consider another representation of cluster analysis. Divjak (2010) represents one of the most complete studies employing multivariate usage-features methodology. In this study, she includes an unrooted or 'phylogenetic' representation of clustering. Assuming that clustering is sound and adequately captures the semasiological structure, this kind of representation is, perhaps, more transparent and interpretable for representing semantic relations. Figure 3, above, is such a phylogenetic graph, produced with a Canberra distance matrix. Just as for a dendrogram, the length of the branches indicates distinctiveness in the clusters. The results mirror those in Figure 2.

2.3 Semasiological clustering with social dimensions

Before we consider the effects of social variation on semantic structure, it must be stressed that one would not expect to find substantial variation with this data and for this lexeme. Therefore, even a small degree of variation is a sign of the extent of the issue. There are four reasons for this:

1. In terms of cultural variation, *run* is a ‘simple’ lexeme. It is the kind of lexeme where one would not expect variation across dialects.
2. In terms of register, *run* is a ‘neutral’ lexeme, not belonging to either formal or informal registers. It is the kind of lexeme where one would expect relatively little variation across text types. One exception to this might be the two central senses of ‘fast pedestrian motion’ versus ‘manage’, where text type would be expected to show variation in use.
3. Although the differences between American and British English are substantial, the dialects remain mutually intelligible for most speakers of both varieties, especially in written language and educated speech. The difference between American and British is not that great, making dialect a good test case.
4. Although there are certainly differences between the registers of spoken conversation and online personal diaries, the style of the latter is extremely informal and is also dialogic. Unlike traditional diaries, authors here engage in discourse with readers and the style of the genre is conversational and casual. Therefore, just as for dialect variation, one would not expect substantial differences in the text type variation.

We could repeat the clustering presented in the previous section for the two dialects and the two registers and compare the clustering. However, the cluster analyses on the full data set are obviously unstable. Halving the data would make any multivariate analysis impossible. Let us begin, rather, with a chi-square test of independence that identifies statistically significant differences along the lines of register and dialect.

A Pearson's Chi-square test of independence for dialect identifies significant differences between the British and American data for the dictionary senses (p = 0.001263). The residuals show that 'become used up' but also 'escape' and 'fast motion' are more typical of the British use, and 'meet', 'increase' but also 'execute' and 'diffuse' of the American use. Register also reveals significant difference (p = 6.376e-05) with the residuals showing that 'escape', 'fast pedestrian motion', 'metaphoric motion' are associated with the diaries, and 'caused motion', 'diffuse', 'execute', 'function', and 'increase' with the conversation data. Having established that there is significant variation, let us move to trying to capture how that variation interacts with the semasiological structure.

Although cluster analysis is a powerful tool for identifying how the different senses are related, it cannot show how register and dialect affect those relations. Ideally, given enough data, we could label the occurrences of the different senses for dialect and register and even both simultaneously. The cluster analysis would then show the relations between the different senses relative to the social factors, clustering, for instance, 'fast pedestrian motion BrEng' and 'fast pedestrian motion AmEng' etc. Although a straightforward procedure, for the number of senses involved, this would require a much larger data set.

Another statistical technique, explained in Glynn (this volume), is correspondence analysis. A multivariate and exploratory technique similar in many ways to cluster analysis, it visualises relations between all the factors considered rather than just one factor. Figure 4 presents the results of a binary correspondence analysis, which examines the interaction of dialect, register, and dictionary sense.

The first two dimensions of the analysis explain 87% of the variation (inertia), which is a relatively stable analysis. Immediately it is visible that American Conversation (AmE.Conv) is distinct in use relative to the dictionary senses, dominating the right two quadrants of the plot on the central axis line. The senses 'increase', 'diffuse', and 'motion into difficulty' (Difficulty) are distinctly and highly associated with the American conversation data point on the right of the plot. In the bottom half of the plot, we find a range of senses distinctly associated with the American diary genre (AmE.Blog). The senses 'campaign', 'copy', and perhaps 'metaphoric motion' (Met.Motion) are highly and distinctly associated with American diary use. 'Meet' and 'extend space' are likely to be associated with American English but not distinct to either register, lying between the two data points for American Diary and American Conversation.

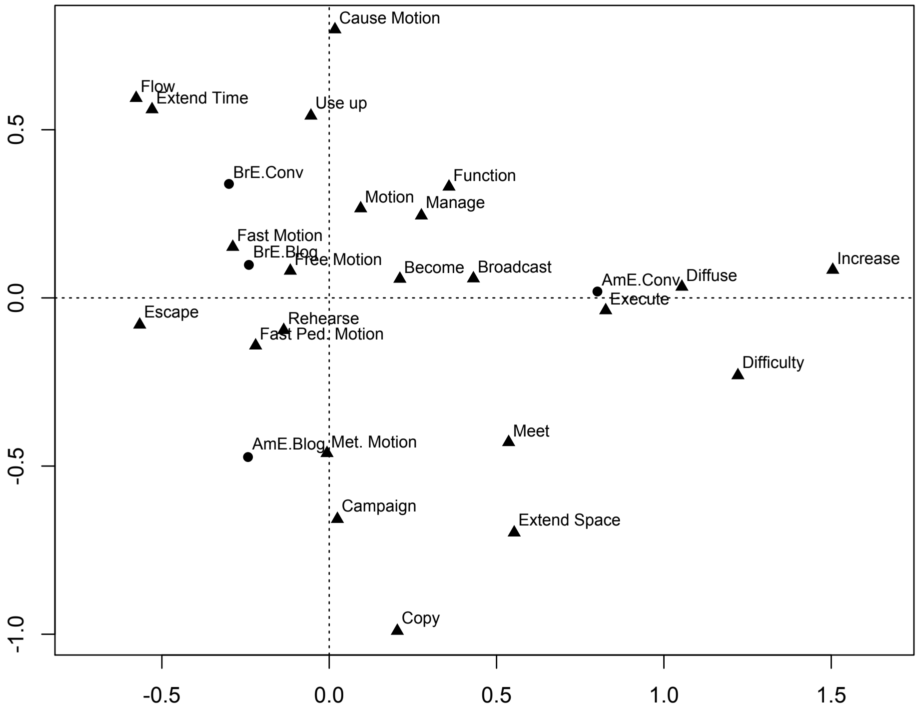

Figure 4. Binary correspondence analysis of register, dialect, dictionary sense

Moving to the British uses, the plot becomes more difficult to interpret. The analysis suggests that there is less register variation in the British sample, the two data points British Conversation (BrE.Conv) and British Diary (BrE.Blog) both lying in the same top left quadrant. Nevertheless, the dialect variation is clear - the senses 'flow' and 'extend time' are highly and distinctly associated with the British use. Other senses, such as 'use up', 'cause motion' and 'escape', are also relatively associated with British use, but this association is not distinctive.

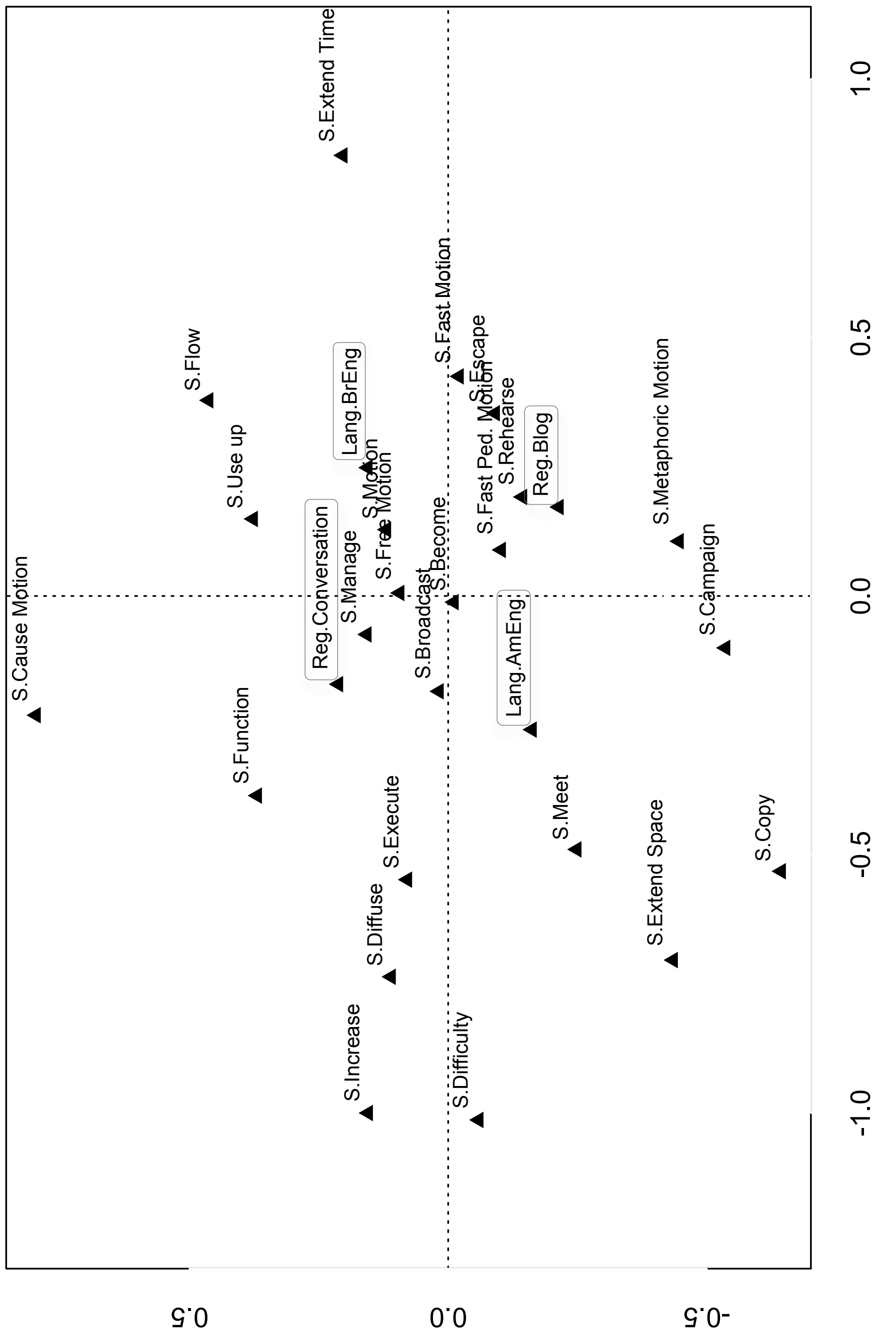

Figure 5. Multiple correspodence analysis
Burt matrix, method 'joint'

In order to obtain a clearer picture of the interactions at hand, let us submit the same data to a multiple correspondence analysis. The binary analysis, in Figure 4, gives us a reliable and stable representation of the associations, but it cannot capture interactions between dialect and register. This is because these two factors were concatenated in order to produce a two-dimensional contingency table for the analysis. We can expand that table into a three-dimensional table and apply multiple correspondence analysis. The results are more difficult to interpret and can be less stable. However, the plot in Figure 5 was produced using the recently developed joint method which addresses both issues of stability and clarity. Fortunately, the results are clear and the explained inertia is 86.7% (Dim. 1: 61.2%, Dim. 2: 25.5%), which for a joint multiple correspondence analysis, using Burt matrices, is a stable result (Greenacre 2007). Further details on and an explanation of the technique of correspondence analysis, its limitations and strengths, can be found in Glynn (this volume, chapter 5).

The results presented in Figure 5 largely reflect the binary correspondence analysis, but by treating the factors of dialect and register independently, the analysis affords us a clearer depiction of their interaction. Each of the four quadrants is characterised by one of the four sociolinguistic features: the top right – British dialect; the bottom right – diary register (Reg.Blog); the bottom left – American dialect; and the top left – conversation register.

We see that senses, such as ‘execute’, and ‘diffuse’ between the American data point and the Conversation data point, are common to these two usage dimensions. The senses ‘campaign’ and ‘metaphoric motion’, lying between the American data point and the register of diary (Reg.Blog), are common to these dimensions. The senses ‘beyond’ the American data point, relative to the British dialect data point in the top right-hand quadrant, are neutral with regard to register, but are distinctly American in contrast to British. These senses include ‘extend in space’, ‘copy’, and ’meet’.

Repeating the interpretation, beginning from the top right-hand quadrant and the British data point, we see that ‘use up’ is distinctly typical of British conversation and that ‘fast motion’ is typical of British diaries. The senses ‘flow’ and ‘extend time’ are less associated with a given register, but are distinctly British, relative to the American data. Again we see that register variation for the British use is less important.

Finally, note the position of ‘manage’ and ‘fast pedestrian motion’. These data points, along with some other senses, are in the centre of the plot. The senses located in the centre are the senses that are not affected by the two sociolinguistic usage factors. These senses are central, but not just in the way that Gries (2006) argued. Although still understood in terms of frequency, we now also have two usage dimensions, dialect and register. Not only are these senses among the most frequent, they are among the senses least affected by context. This is a crucial refinement to the frequency operationalisation of (proto)typicality – uses that that are common (frequent) across all contexts are more central to the meaning of a lexeme. This finding is equally as important as discerning which senses are typical of specific contexts.

Gries (2006) stresses that the small sample means that the study can only be seen as a methodological test, rather than a fully descriptive analysis. For these reasons, the statistical techniques employed are only exploratory. He suggests the use of configurational frequency analysis to identify statistical significance in the results, allowing one to determine which correlations are not chance, and which may be simply a result of the small sample. Although configurational frequency analysis would be an excellent choice for this, it requires more data than is available in either study. It also follows that, with more data, log-linear regression or polychotomous logistic regression would be even better, giving not only statistical significance but also predictive strength to the model. Such analyses are now within the capabilities of corpus-driven research, but require a larger scale analysis.

Moreover, before such an analysis is undertaken, the identification of senses must be better operationalised. The analysis of the usage-features must be found to cluster into senses and then these multivariate senses must be shown to be statistically significant. With senses based on clusters of usage-features (ID Profiles), rather than revealed by matching occurrences with dictionary entries, we can then return to the clustering. This step in corpus-driven polysemy research has begun (Glynn 2009, 2010a, in press), but remains at the initial stages. Once we are armed with the analytical tools to identify multivariate senses (rather than dictionary senses), then we need to progress to modelling the semasiological structure and the prototype effects, using more advanced statistical procedures such as configurational frequency analysis and log-linear regression. The present purposes are to demonstrate that sociolinguistic effects must be integrated into the study of prototype structuring. To these ends, let us submit the data to multiple logistic regression.

Explained in Speelman (this volume), logistic regression is a confirmatory multivariate technique that allows us not only to determine which of the usage features and / or dictionary senses are significantly associated with either of the sociolinguistic factors, but it also enables us to determine how important that association is.

Logistic Regression - Dialect

Let us begin with dialect. Three logistic regression models are reported: a multiple model based on usage-features excluding dictionary senses (Model 1); a second multiple model that includes dictionary senses (Model 2); and a simple model with the dictionary senses as a sole predictor variable (Model 3).

Table 2. Logistic Regression Models for Dialect

		Model 1	Model 2	Model 3
Coefficients	Transitivity - Transitive	0.596619*	0.487405°	-
	Tense - Past	0.658282*	-	-
	Tense - Present	0.387003	-	-
	Aspect - Progressive	-	0.178450	-
	Aspect - Simple	-	0.542254°	-
	Mood - Imperative	0.533600	0.392821	-
	Mood - Interrogative	1.048808*	1.094234*	-
	Clause Type - SubPronoun	-1.807732°	-2.225015	-
	Clause Type - SubNP	-0.490190	-2.225015°	-
	Clause Type …	…	…	-
	Subject - Human	-1.446052°	-	-
	Subject - Locations	-0.289667	-	-
	Subject - Machine	-1.564679*	-	-
	Subject …	…	-	-
	Sense - Use Up	-	0.955200°	0.94852*
	Sense - Diffuse	-	-1.364175°	-1.65945*
	Sense - Execute	-	-1.196454*	-1.14862*
	Sense - Campaign	-	-1.289732	-1.65945
	Sense - Fast Motion	-	0.792536	0.82546
	Sense - Flow	-	1.405831	1.55943
	Sense - Increase	-	-2.287706*	-2.12945*
	Sense - Meet	-	-1.566828*	-1.59046*
	Sense …	-	…	…
Model Statistics	d.f.	20	27	20
	G^2	41.39**	70.71***	60.16***
	ROC	0.668	0.716	0.671
	Nagelkerke R^2	0.112	0.186	0.156
	Bootstrapped R^2	0.0138	0.0435	0.057

The models are all checked for multicolinearity, and factors producing a variance inflation of more than 2.5 are removed.[4] Moreover, the models are checked for singularity with a Kappa calculated condition number - any model with a value higher than 6 is rejected.[5] The strict check on variance inflation and singularity assure an orthogonal model. The models are also checked for influential observations as well as overfitting, neither of which is a problem. Outliers are not removed. In a backward elimination of factors, model selection was based on significance values and the Akaike's information criterion (AIC), not on predictive strength.[6] For readers unfamiliar with logistic regression, the testing of the model and criteria for acceptability were extremely strict, making the results as conservative as possible.

For the sake of brevity, some non-significant levels are omitted, indicated by '…'. Positive coefficients predict British English and negative coefficients ("-") predict American English. Since we are comparing models, only the coefficients and some essential model statistics are reported.[7] In Table 2, the coefficients for each of the levels (usage-features) are listed with the alpha levels (° $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).[8] The degrees of freedom (d.f.), Log Likelihood chi-squared or deviance measure (G^2), the C index or coefficient of concordance (C), the Nagelkerke pseudo R^2 (N. R^2) and a bootstrapped estimation of the pseudo R^2 (Boot. R^2) are included as model statistics.

Models 1 and 2 are the result of a backward elimination of a full model without and with dictionary senses respectively. Object and complement usage-features were, interestingly, found to be not at all significant. This could be due to the large amount of features (levels) belonging to these factors. A much larger data set is needed to handle such complexity. For these same reasons, it was not possible to examine interactions. In order to improve the model, some small cells were removed for the variable of dictionary sense. This weakens predictive strength, but improves the AIC and overall parsimony of the model. In model two, entering subject semantics and dictionary senses simultaneously shows signs of multicolinearity. The subject is, thus, omitted.

Although models 1 and 2 show some significant features, we are not interested in finding differences between the two dialects, but differences in the semasiological structuring of the two dialects. Since we do not know how the different formal usage-features interact in terms of semasiological structure, it is difficult to interpret models 1 and 2 in these terms. It is clear, however, that these usage-features do not offer a coherent or predictive model, which reassures us that there are no obvious differences that could superficially distinguish one dialect from the other.

Model 3, on the other hand, is simple to interpret. If we recall the residuals of the Chi-square test shown above, we see confirmation of those results, but this time with statistical significance as well as a score indicating degree of effect. For the American data, the sense 'increase' is by far the most distinctive feature, followed by 'execute', 'meet' and then 'diffuse'. 'Campaign' is not statistically significant (*p*-0.1338), but it would almost surely become significant with more data. We saw in the correspondence analysis that 'campaign' is highly associated with American usage and it is known that this sense is effectively unique to the American dialect. In British English, the verb *stand for* an elected post is more typical than *run for* an elected post. The sense 'motion into difficulty' behaved similarly to 'campaign', but was removed due to its small count.

With the current data set, five of the senses are statistically significant. This confirms what we saw in the Chi-square test and correspondence analysis. In the logistic regression, the coefficients give us a rank of influence similar to the Pearson's residuals obtained from the Chi-square tests. This kind of ranking is exactly the type of information needed for understanding the effects of such dialect variation on prototype structuring.

However, none of the five senses in question is a particularly strong predictor, though the sense 'increase', associated with American usage with a coefficient of 2.1, is quite strong. At the other end of the spectrum, 'become used up' predicting British English is a relatively weak predictor. Ranked in order of influence, we now know that 'increase', 'diffuse', 'meet' and then 'execute' are distinctly American in use, where only 'become used up' is a significant predictor for British usage.

The statistics presented beneath the table of coefficients show that the model is not in the least predictive. A fourth model, without the infrequent senses of 'campaign' and 'motion to difficulty', produces comparable statistics (G^2: 57.18, d.f. 18, R^2: 0.151, C: 0.667). The poor predictive strength of the model, of course, is to be expected. If the differences between the dictionary senses in themselves were so great that one could predict one dialect over the other with this information alone, we would have such obvious semasiological variation, that this study would not be needed. What the logistic regression gives us is a clear and specific picture that although all the senses are possible in both dialects ('campaign' aside), the differences in frequency of occurrence are great enough that even with a small data set, significant differences can be identified.

Register Effects

Just as for the dialect effects, we will consider three models, a single regression analysis of the dictionary senses and two multiple logistic regressions. Table 3 summarises the models. Model selection followed the same criteria as for the previous logistic regression. Positive coefficients predict conversation register and negative coefficients predict diary register.

Table 3 Logistic Regression Models for Register

Usage-feature (levels)	Model 1	Model 2	Model 3
Tense - Past	-0.57321º	-	-
Tense - Present	0.0518	-	-
Clause Type - Subj. NSub	-	-1.12794*	-
Clause Type - Subj. Pronoun	-	-0.46960	-
Clause Type - Subj. NP	-	-0.67469*	-
Subject Sem. - Human	1.21485º	-	-
Subject Sem. - Location	2.30656*	-	-
Subject Sem. - Machine	2.53738**	-	-
Subject Sem. - Quality	2.56357*	-	-
Subject Sem. …	…	-	-
Object Sem. - Animate	2.61971 +	-	-
Object Sem. - Concrete Count Noun	1.92852**	-	-
Object Sem. - Concrete Mass Noun	2.12490**	-	-
Object Sem. - Events	2.09651 **	-	-
Object Sem. - Human	1.60972*	-	-
Object Sem. - Location	1.68428**	-	-
Object Sem. - Machine	2.26511**	-	-
Object Sem. - Organisation	2.30594	-	-
Object Sem. - Quantity	1.65779º	-	-
Object Sem. - NA	1.17705º	-	-
Object Sem. …	…	-	-
Sense - BecomeUsedUp	-	1.12163*	0.9639*
Sense - Diffuse	-	2.09507**	1.9373*
Sense - Escape	-	-0.60738º	-0.4453
Sense - Execute	-	1.43060*	1.4265**
Sense - Function	-	1.92870**	1.8684**
Sense - In Charge of	-	1.36388*	1.1164 *
Sense - Increase	-	2.77507*	2.4073*
Sense - Manage	-	0.82270º	0.5691
Sense - Meet	-	0.73929º	0.4457
Sense - Motion to Difficulty	-	1.88122*	1.8320*
d.f.	25	26	19
G^2	67.45***	82.31***	63.95***
ROC	0.697	0.721	0.69
Nagelkerke R^2	0.176	0.211	0.167
Bootstrapped R^2	0.064	0.077	0.063

Just as for the regression analyses above, none of the models are predicatively strong. This means that, for the features analysed, we cannot predict whether an example will be one dialect or the other or one register or the other. This does not mean, however, that we cannot interpret the table of coefficients to see where significant differences do exist.

In model 1, we have an interesting selection of significant Subject Semantic and Object Semantic categories. The interaction of such features is likely to represent the usage-configurations of profiles that could be understood as non-reified senses. Evidence for this can be found in high colinearity that is produced when these variables entered into the regression with dictionary senses. Of course, further research is needed to ascertain of what these configurations would consist.

Many individual (as opposed to configurations of) Subject and Object Semantic categories could be seen as operationalisations of different dictionary senses in their own right. For example, the Object Semantic feature of 'machine' would be a reasonable operationalisation of the sense 'operate' (or 'execute'), just as the Subject Semantic category of 'machine' would indicate 'function'. We see here highly significant and important predictors of the register conversation. This is the kind of extra-linguistic effect on frequency-based prototype structure we referred to in section 1, when we compared *run* - 'manage' and *run* 'fast pedestrian motion'. Rather than dictionary senses, we see how different semantic features are interacting with other dimensions of use. Although not predicatively strong, model 1 offers interesting insights into the kind of semantic variation we have between the two registers.

Models 2 and 3 differ little. The addition of other factors in the multiple regression of model 2 merely identifies some syntactic variation. The dictionary senses again confirm what we saw in the chi-square test, although we now see that none of the senses is significantly associated with the diary register. We see also that the senses 'increase', 'motion to difficulty' and 'diffuse' are the senses most highly associated with conversation.

3. Summary

This study examined the semasiological variation of *run*, replicating the study of Gries (2006), but adding two sociolinguistic contexts. The aim was to add two usage dimensions to the polysemy 'map' in order to more accurately represent usage-structure. The descriptive findings of Gries are largely confirmed. One further sense, 'escape', was found to be relatively important. More importantly, the study added to Gries' results by identifying which of the non-central senses appear to be sociolinguistically varied and also by discerning which senses are neutral with regard to the two contexts considered. That certain senses are not affected by sociolinguistic variation, yet others are, adds considerable weight to the argument that they represent the prototype senses.

The theoretical and methodological goal of this study was straightforward – to demonstrate that, although at times subtle, sociolinguistic factors have a significant impact upon semasiological structure. Cognitive Linguistics has propounded a usage-based model of language since its beginning, but its approach to semantic structure remained largely Structuralist and Generativist in its assumptions and methodologies. The increased use of observational techniques as well as multivariate statistics improves our understanding of the complexity of polysemic structures, but also brings out the need to treat semantic structure in a radically new way. Our very own model of language states that language structure is varied and emergent, that categorisation is rarely discrete. We must accept the ramifications of this in our research and resist the temptation to assume that discrete reified lexical senses exist or that those senses exist in some abstract system, independent from the variation of societies and cultures that use them. Since meaning is emergent, multidimensional, and ultimately non-reifiable, a description of polysemy that is both cognitively and communicatively realistic will depend upon developing empirical methodology that can adequately describe the complexity of this object of study.

Notes

1. Lakoff (1987) was an early protagonist of both the theory of prototype categorization and the model of radial network analysis. Prototype theory was developed and refined by Geeraerts (1989, 1993, 1997), Taylor (1989), and Kleiber (1990). Radial network analysis was developed and formalised by especially Rudzka-Ostyn (1989), Cuyckens (1993), and Janda (1993).

2. Kudrnáčová (2010) has also followed up Gries' (2006) study with a more fine-grained corpus-based semantic analysis. Her study is not quantitative, but her corpus-based insights will inform future research. In descriptive terms, the next step is to apply a more detailed usage-feature analysis and begin, not with dictionary senses, but a range of subtle semantic features. The senses should then be clusterings of those semantic features rather than simply matches between dictionary entries and observed occurrences.

3. Divjak & Gries (2006: 37) state that the Canberra distance matrix is best suited to small cell counts, such as we have here. Gries (2009: 317) says the choice is subjective. Gries & Stefanowitsch (2010: 79) employ the Manhattan distance matrix, citing Levy *et al.* (1999) as justification. Levy *et al.*'s study compares five distance matrices but not Canberra. It seems that the question of how the choice of distance matrix affects the results needs to be investigated systematically.

4. Some authorities indicate a variance inflation factor of 10 to be acceptable (DeMaris 2003: 517; Dodge 2008: 96; Chatterjee & Hadi 2006: 238; Marques de Sá 2007: 307; Speelman p.c.), other authorities are non-committal (Faraway 2002: 117-120; Maindonald & Braun 2003: 201-3). Glynn (2010) and Speelman (this volume) opt for a maximum inflation value of 4. Szmrecsanyi (2006: 215) notes that even values as low as 2.5 can be a cause for concern. Multicolinearity is a serious issue in regression and can lead to Type I errors. Since we do not necessarily understand the relationship between many of the factors in our model, we will opt for a maximum VIF score of 2.5.

5. Baayen (2008: 182) states that a condition number between 0 and 6 indicates no multicolinearity and 15 indicates a medium degree.

6. The AIC score is a score that helps compare the parsimony of different models. The scores are relative and a lower number indicates a more parsimonious model.

7. The R output includes the estimated standard errors and the Wald Chi square, or *z*-test (z) obtained by dividing the coefficient by its error. See Speelman (this volume) for a detailed explanation. The output of lrm is explained in Baayen (2008: 2004) and Gries (2009: 297). See also Chatterjee & Hadi (2006).

8. Significance levels are primarily used for model selection, assuming a level is significant, caution should be taken in interpreting them relatively (Faraway 2002: 126).

References

Arppe, Antti & Juhani Järvikivi. 2007. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3: 131–159.

Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert & Arne Zeschel. 2010. Cognitive corpus linguistics: Five points of debate on current theory and methodology. *Corpora* 5: 1-27.

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Chatterjee, Samprit & Ali Hadi. 2006. *Regression Analysis by Example*. Chichester / New Jersey: Wiley.

Cuyckens, Hubert. 1993. The Dutch spatial preposition “in”: A cognitive-semantic analysis. In C. Zelinsky-Wibbelt (ed.). *The Semantics of Prepositions*. *From mental processing to natural language processing*, 27-72. Berlin: Mouton de Gruyter.

DeMaris, Alfred. 2003. *Regression with Social Data: Modeling continuous and limited response variables*. Chichester / New Jersey: Wiley.

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. 2010. *Structuring the Lexicon: A clustered model for near-synonymy*. Berlin / New York: Mouton de Gruyter.

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

Dodge, Yadolah. 2008. *The Concise Encyclopedia of Statistics*. Heidelberg / New York: Springer.

Faraway, Julian. 2002. *Extending the Linear Model with R: Generalized linear, mixed effects and nonparametric regression models*. London: Chapman and Hall.

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

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

Geeraerts, Dirk. 1987. 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. 1993. Vagueness’s puzzles, polysemy’s vagaries. *Cognitive Linguistics* 4: 223-272.

Geeraerts, Dirk. 1997. *Diachronic Prototype Semantics. A contribution to historical lexicology*. Oxford: Clarendon Press.

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

Gilquin, Gaëtanelle. 2010. *Corpus, Cognition and Causative Constructions*. Amsterdam / Philedelphia: John Benjamins.

Glynn, Dylan. 2009. Polysemy, syntax, and variation. A usage-based method for Cognitive Semantics. In V. Evans & S. Pourcel (eds.), *New Directions in Cognitive Linguistics*, 77-106. Amsterdam / Philedelphia: John Benjamins.

Glynn, Dylan. 2010. Testing the hypothesis. Objectivity and verification in usage-based Cognitive Semantics. In D. Glynn & K. Fischer (eds.), *Quantitative Cognitive Semantics. Corpus-driven approaches*, 239-270. Berlin / New York: Mouton de Gruyter.

Glynn, Dylan. In press. Cognitive socio-semantics. The theoretical and analytical role of context in meaning. *Review of Cognitive Linguistics*.

Gries, St. Th. 2003. *Multifactorial Analysis in Corpus Linguistics: A study of particle placement.* London / New York: Continuum Press.

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

Gries, Stefen Th. 2009. *Statistics for Linguistics with R: A pratical introduction*. Berlin: Mouton de Gruyter

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

Kleiber, George. 1990. *Sémantique du prototype : catégorie et sens lexical*. Paris: Presses Universitaires de France.

Kudrnáčová, Naděža. 2010. On pragmatic and cognitive processes in meaning variation. *Linguistica Silesiana* 31: 55-67.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things. What categories reveal about the mind.* Chicago / London: University of Chicago Press.

Lakoff, George. 1990. The Invariance Hypothesis. *Cognitive Linguistics* 1: 39-74.

Lehrer, Adrienne. 1982. *Wine and Conversation*. Bloomington: Indiana University Press.

Lehrer, Keith & Lehrer, Adrienne. 1994. Fields, networks, and vectors. In F. Palmer (ed.), *Grammar and Meaning. A Festschrift for John Lyons*, 26-47. Cambridge: Cambridge University Press.

Levy, Joe, John Bullinaria, & Malti Patel. 1999. Explorations in the derivation of-cooccurrence statistics. *South Pacific Journal of Psychology* 10: 99-111.

Maindonald, John & Braun, John. 2003. *Data Analysis and Graphics Using R: An example-based approach*. Cambridge. Cambridge University Press.

Marques de Sá, Joaquim. 2007. *Applied Statistics Using SPSS, STATISTICA, MATLAB and R.* Heidelberg / New York: Springer.

Rosch, Eleanor. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology* 104: 192-233.

Rudzka-Ostyn, Brygida. 1989. Prototypes, schemas, and cross-category correspondences: The case of *ask*. In Dirk Geeraerts (ed.), *Prospects and Problems of*

Prototype Theory (Special ed. *Linguistics 27*), 613-661. Berlin / New York: Mouton de Gruyter.

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

Schinka, Jane & Velicer, Wayne (eds.). 2003. *Research Methods in Psychology*. Vol. 2. *Handbook of Psychology*. New Jersey / Chichester: Wiley.

Schmid, Hans-Jörg. 2000. *English Abstract Nouns as Conceptual Shells. From corpus to cognition*. Berlin / New York: Mouton de Gruyter.

Saussure, Ferdinand de. 1997 [1916]. *Cours de linguistique générale*. Lonrai : Normandie Roto.

Szmrecsanyi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin / New York: Mouton de Gruyter.

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

Tribushinina, Elena. 2009. On prototypicality of dimensional adjectives. In J. Zlatev, M. Andrén, M. Johansson Falck & C. Lundmark (eds.), *Studies in Language and Cognition*, 111-128. Newcastle: Cambridge Scholars.

Tyler, Andrea & Evans, Vyvyan. 2001. Reconsidering prepositional polysemy networks: The case of *over*. *Language* 77: 724-65.