

УДК 81'373.612

CORPUS-DRIVEN COGNITIVE LINGUISTICS. A CASE STUDY IN POLYSEMY

Dylan GLYNN

*Senior research fellow
Centre for Language and Literature
Lund University, Sweden EU*

This study introduces quantitative corpus methods for Cognitive Linguistics whose usage-based model permits conceptual analysis based on corpus data. By assuming that grammar is a result of repeated use (entrenchment), one can examine patterns of use in large amounts of language in order to 'map' its grammar (conceptual structure). Examining the English lexeme *hassle*, the case study presents one method for such analysis and applies it to the interaction of morpho-syntactic and lexical semantics. The analysis employs two statistical techniques that aid in the search for and the verification of usage-patterns in language.

Key words: cognitive linguistics, quantitative corpus methods, conceptual analysis.

Introduction. The 21st century is witnessing an empirical revolution in linguistics. Drawing on 'found' data (corpora) or elicited data (experimentation), linguists are leaving behind the theoretical debates of the previous era and turning to natural language data, proposing falsifiable hypotheses, and testing them with empirical methods. Cognitive Linguistics, as a theoretical approach, is one of the best-placed paradigms to inform this revolution; indeed, it arguably was integral to its onset. This study examines a quantitative corpus-driven method within the theoretical framework of Cognitive Linguistics.

The end of the 20C saw the theoretical paradigm of Cognitive Linguistics [40, 38] dominate the international linguistic research community. Its theoretical approach is a quasi-descendant of British and Prague School Functionalism but adopts much, despite itself, from Structuralism [44] and Mentalist-Modularism [3]. Indeed, for many years, Cognitive Linguistics defined itself in contrast to these previous paradigms. However, despite its success in revealing the shortcomings of the previous approaches, Cognitive Linguistics has struggled to offer a viable alternative. The end of the century witnessed a realisation within the cognitive research community that although Structuralism and Mentalist-Modularism had been adequately refuted, Cognitive Linguistics itself had proposed an inadequate alternative. This short discussion will chart one of the most important responses to this inadequacy from within the community.

1. Theoretical Tenets of Cognitive Linguistics. The theoretical tenets of Cognitive Linguistics are straightforward. The theory holds that the failures of 20C linguistics lie in a single assumption: one can examine and understand the functioning of 'parts' of language separately. This assumption was tempting because it is reasonable. It is reasonable to divide a large and difficult task into different parts and to attempt that task piece by piece. The vast complexity of language is surely such a task and when de Saussure distinguished the *langue* – *parole*, paradigmatic – syntagmatic, and the diachronic – synchronic dimensions of language, he did so in an attempt to render language more rigorously analysable. Chomsky followed suite. He returned the place of the individual mind to the model but

removed meaning, proposing a syntactic model that generated language. His principle was the same – cordon off parts of language to permit rigorous analysis. Cognitive Linguistics can be summarised as an approach that argues this separation of language into structural parts is reductionist and will never produce adequate results. In brief, Cognitive Linguistics holds that no matter how varied and complex language be, we must approach it holistically.

What exactly does a holistic approach to language entail? Firstly, in contrast to Structuralism, there is no distinction between semantics and pragmatics. Meaning is understood as all that one knows about the world [9, 38]. This ‘encyclopaedic semantics’ is foundational to Cognitive Linguistics and its impact upon analysis cannot be underestimated. To appreciate the importance of the assumption, it can be rephrased – meaning is how people use words in context. This rephrasing will become essential below, because it permits an operationalisation of meaning in empirical terms. Secondly, in contrast to Mentalist-Modularist theories of language, a holistic approach means that there is no distinction between syntax and lexis – all linguistic forms, regardless of their schematicity, integrate and are used relative to each other [40]. This can also be rephrased – lexical semantics must combine felicitously with the syntactic semantics but also with the morphological semantics and the prosodic semantics, and all simultaneously. From such a perspective, the different degrees of formal schematicity (such as lexis, syntax, prosody etc.) are epiphenomenal. Any effort to distinguish syntax from other formal structures is fruitless because its use is entirely integrated into a complex system. Highly schematic constructions, such as the transitive construction, must integrate with less schematic constructions, such as the caused-motion construction, and then this with tense, mood, *Aktionsart*, intonation and the referential semantics of the lexemes. Moreover, the meaning of all of the above is not restricted to linguistic semantics in Structuralist terms, but full encyclopaedic semantics including register, gender, dialect and so forth.

Having established (or re-established) the full complexity of language, Cognitive Linguistics proposes two basic analytical constructs to enable generalisations to be made about this dynamic and complex phenomenon. The first analytical assumption concerns grammaticality and is termed ‘entrenchment’. This notion can be understood as the ‘fixing’ of linguistic knowledge through repeated use and it is basic to the Usage-Based Model of grammaticality [41, 42]. In this language model there is no *langue* in the de Saussurian sense and there is no ‘linguistic competence’ in the ‘ideal speaker’ sense of Chomsky. The Usage-Based Model is simple: the more often a speaker is exposed to a form-meaning pair, the more often it is automated as part of the language system for that speaker. It follows that each individual has his or her internal knowledge of a language. A grammar is, therefore, a generalisation over those ‘knowledges’ of many speakers. To the extent that speakers share this knowledge of the use of a linguistic form(-meaning pair), it is ‘grammatical’. Form-meaning pairs that are less commonly shared are less grammatical. This is often operationalised in terms of frequency of use – the more often or the more rarely a form-meaning pair occurs, the more grammatical or less grammatical it is. However, that frequency alone can account for entrenchment has been brought into question [1, 20].

The second analytical assumption concerns semantic structure and entails a theory of categorisation. Although it is held that meaning includes all that we know of the world, this does not entail that this knowledge is a sea of unorganised or unrelated concepts. It is assumed that connections between concepts exist, that all perceived and conceived things are

understood as similar and dissimilar to things already experienced. This categorisation is referred to as Prototype Set Theory [38].

Geeraerts [10] and Taylor [31] offer a complete summary of the issues involved in categorisation and Glynn (in press b) offers an explanation of how these theories are operationalised in corpus-driven research. Essentially two categorisation effects are assumed to account for how we determine similarity and dissimilarity. Prototype categorisation is evoked for non-scalar concepts that possess good and less good examples of reference (e.g.: *bird*, *sparrow*, and *penguin*) and fuzzy categorisation for scalar concepts that possess gradable referents (e.g.: *short* — *tall*). Importantly, such categorisation is not restricted to referential concepts or lexical semantics but is equally valid for grammatical semantics and also generalised cultural concepts called Idealised Cognitive Models [38, 39]¹.

Although most linguists today would agree with most of the discussion above, this does not mean that Cognitive Linguistics has solved the mysteries of language, indeed far from it. In the 1990s, it became apparent that although such a model of language maybe theoretically sound, it is difficult to produce falsifiable hypotheses based upon it. In a system where all form and meaning and context are relevant and all grammaticality is relative, how can one test the generalisations one makes about language structure? Language is a complex, dynamic, and varied system – agreed. However, this does not mean that, as scientists, we can make claims about its structure without a method for disproving those claims. This was made apparent in an important experimental study ‘Mirroring whose mind, the language user’s or the linguist’s’ (Sandra & Rice’s 1995). It was shown that the kind of semantic analysis propounded by Lakoff [38] and Cuyckens [5] *inter alia*, was *ad hoc* – different analyses, by different linguistics, would produce different results. Although semantic analysis is necessarily subjective, this must be adequately operationalised to permit repeat analysis and hypothesis falsification. Rather than a crisis within the cognitive community, this brought about a race to develop methods and analytical frameworks that can falsify results. Using both psycholinguistic experimentation and corpus-driven analysis, the beginning of the 21st century has seen a veritable flood of research that employs the functional-cognitive model of language but proceeds empirically, seeking to test hypotheses inductively rather relying uniquely on intuition. This does not mean that intuition is not needed, indeed it is crucial, but one must find ways of operationalising subjective analysis and testing the descriptive adequacy of results.

2. Corpus-driven Cognitive Linguistics. A full description of corpus driven research in Cognitive Linguistics research is beyond the means of this discussion. Tummers & al. [52] and Gilquin & Gries [13] offer excellent discussions on the state of the art. Glynn [17] summarises the field as it stands. However, in order to appreciate how corpus results can be used to understand semantic, and indeed conceptual structure, we need to briefly discuss the operationalisation of the analytical assumptions of Cognitive Linguistics. Operationalisation is the process of defining a scientific concept or object of study in a way that makes it measurable in terms of empirical observations. This approach to defining

¹ The notion of the usage-based model is far from a uniquely Langackerian or even cognitive one. Many theoretical approaches, especially within functional linguistics, assume this model of language. Culioli [4] and Givón [16] are examples at hand.

analytical concepts has slowly become standard in the social sciences, but is still sometimes ill-understood in linguistics.

Operationalisation of conceptual structure is the basic principle behind corpus-driven Cognitive Linguistics. Corpus analysis identifies re-occurring combinations (patterns) of linguistic forms, meanings, and contexts. These patterns of language use are argued to reflect a speaker's internal knowledge of language. Since, in Cognitive Linguistics, we believe grammar, and language generally, to be conceptually motivated, language use not only mirrors knowledge, but also conceptual structure. This operationalisation permits two things: (i) a direct method for making generalisations across large numbers of speakers (thus a language's grammar) and (ii) an indirect method for producing hypotheses about the conceptual structure of a language (motivation for a language's grammar). Let us consider these two possibilities. If we assume that language use equates language structure and if we capture patterns of use by large numbers of native speakers, then we are making generalisations about a language, in other words, its grammar in the fullest sense of the word, including lexical relations and so forth (remember Langacker's idea of integration and grammaticality). Moreover, if we have the context and the reference of this language use, then we can make generalisations about how that is used, to what ends. This conceptual-functional dimension permits the proposal of hypotheses about semantic structure, again in the fullest sense of the word (remembering Lakoff's encyclopaedic semantics and categorisation). These principles are assumed in corpus-driven cognitive research. Glynn (in press, *forthc.* a) offers further discussion on the matter, but the principle is implicit in all corpus-driven Cognitive Linguistic research.²

Analytically, two approaches are popular within the field. The first relies exclusively on formal patterns of usage. This tradition grows out of the British functionalist corpus linguistics [37, 47]. These exclusively formal corpus techniques, such as collostructional analysis [48, 49, 34], examine collocations and, using tests for statistical significance, look for patterns of formal association. These formal associations are then interpreted semantically. A second approach extracts large samples of a given form in context, and then examines its formal, semantic, and extralinguistic use. This approach is referred to as the profile-based or usage-feature approach [33, 30, 8, 6, 32, 55, 19, 21, 22, 36]. The profile-based approach has the advantage that full multifactorial statistics can be applied as well the advantage that more subtle semantic factors can be captured. However, it has two disadvantages. The first disadvantage is that semantic analysis necessarily weakens the objectivity. The counter argument to this is that one may only progress so far in the description of language without considering semantics and that ultimately, one must move towards subjective analysis. The key is to operationalise the semantic analysis sufficiently to enable repeat analysis. The second disadvantage is that manual analysis is labour intensive, restricting considerably the number of examples in the data set. This, in turn, reduces the representativity of the

² This should not suggest that Lakoff is unique in the study of culturally determined cultural concepts. However, his formalisation of the approach is the most well-known. The Slavic linguistic communities have a rich tradition in this field; important examples would include Wierzbicka [53], Bartmiński [2], and Степанов [59], Кубрякова & Янко [58] and Воркачев [57] *inter alia*. Extending the methods presented in this study to culturally broad concepts is the next and evident step. Glynn [26] is representative of such development.

sample. The only response to this is that testing for statistical significance becomes all the more important.

In order to understand how these assumptions work in linguistic analysis, let us turn to a small sample of an analysis. The analysis is taken from Glynn (forthc. a.) and is typical of the kind of research presented in the profile-based approach.

3. Lexical Semantic Relations. Capturing the integration of grammatical and lexical semantics. According to the model of language propounded by Cognitive Linguistics, all variation in use is semantic variation and therefore, polysemy. This section presents a 'mini' case study of the polysemy of the lexeme *hassle*. Remembering the task is to capture semantic structure with a holistic approach to language, we must account for the variation in use of the lexeme in context, both linguistic context and social context. This study will show one method for capturing this multidimensional aspect of meaning. For practical purposes, it restricts itself to syntactic context, or, in other words, the interaction of morpho-syntactic and lexical semantics.

The data consist of 700 observations, proportional to the various forms of the lemma, taken from British and American on-line personal diaries. These examples were manually analysed for a wide range of formal, semantic, and sociolinguistic usage-features. Altogether, 120 features across 20 factors were annotated (indexed in a tabular database). Limits on space prevent the presentation of this analysis, but let us consider four of the more difficult and subjective factors: the type of the Agent; the Cause of the event; the kind of Affect experienced by the Patient; and the use of Humour in the utterance. Each of these factors is made up of a range of usage-features such as 'anger', 'boredom', 'interruption' or 'emotional pain' and so forth. These features, like the features of a phonological analysis, combine in different ways to reveal usage patterns. Due to their highly subjective nature, these three factors were each coded independently by two linguists. The Cohen's Kappa is used to calculate the degree of inter-coder agreement. The rules of thumb over agreement rating follow: Strength of agreement < 0.2 Poor; > 0.2 – 0.4 Fair; > 0.4 – 0.6 Moderate; > 0.6 – 0.8 Good; > 0.8 – 1 Very good. Despite the fact that this calculation is considered a conservative measure, the inter-coder agreement was approaching 1, which is perfect.

Affect – $k = 0.949181$

Cause – $k = 0.9510682$

Humour – $k = 0.9110852$

These figures show us that although the features in question are highly subjective, they were sufficiently well operationalised that repeat analysis gave near-identical results.

Another important factor is that of Syntax. Although many syntactic patterns were identified, at a relatively coarse grain level of analysis, eight basic morpho-syntactic profiles were found.

- (1) a. Transitive: The awful stalker guy who's been hassling me
- b. Transitive Oblique: If you hassle me about my kinky hair, I'll cut it all off.
- c. Intransitive: Officer McCoy, me and him was hassling and my gun went off
- d. Noun Mass: It saves all that ammoying hassle of SOD'S-BLOODY-LAW!!!!!!
- e. Nominal Count: I rarely paint my nails(It can be such a hassle!)
- f. Gerund: the technical know-how to do this sort of hassling
- g. Adjective Attributive: It's a very hassily event to do.
- h. Adjective Predicative: She will not take part, she is tired and hassled

3.1 Identifying patterns of usage

Other factors analysed include tense, aspect, polarity, person, number, as well as range of semantic features such as the telicity, plexity, concreteness, animacy, and boundedness of the arguments in the event. This analysis results in large tables of features, listed for each example. Although one can look at relative frequencies in such a table, it is effectively impossible to identify patterns therein, especially if one is interested in the interaction of many usage-factors. In order to identify usage patterns across such large tables, there exist a range of statistical tools. In this study, we employ multiple correspondence analysis. This technique calculates the frequency of co-occurrence of features and converts these frequencies to spatial distances. If we examine, for instance, four factors, the relationship in use between the features will be expressed as relative distances in a four-dimensional space. The technique flattens this to a 2-dimensional space, which it plots. Therefore, in the plots, proximity represents correlation in use – data points close to each other are highly associated usage-features and data points far from each other represent dissociated usage-features.

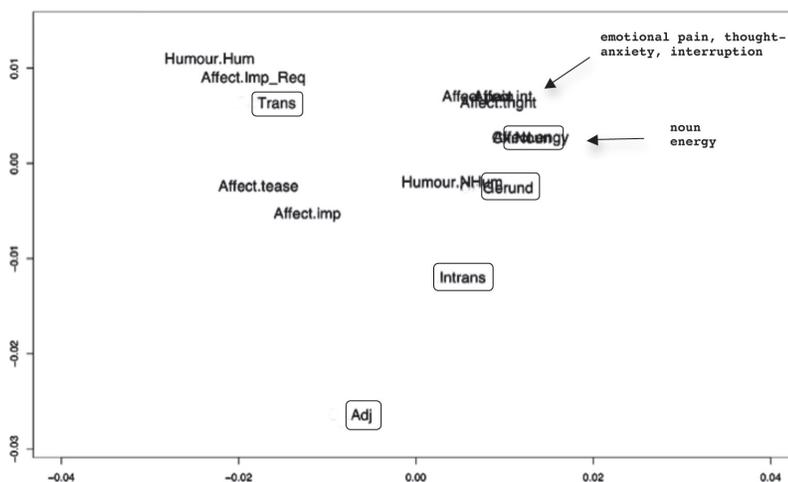


Fig. 1. Interaction of Syntax, Cause, Affect, and Humour
Multiple Correspondence Analysis (Euclidean distance Matrix)

Figure 1 shows the results of correspondence analysis of the factors of Syntax, Cause, Affect, and Humour. Although difficult to read, two distinct usage clusters emerge. On the right, the clustering of usage-features is so tight that it becomes illegible. Labels have been added to help the reader identify the features in question. We see the Noun <Noun> and Gerund <Gerund> forms are associated with a range of Causes and Affects, such as ‘emotional pain’ <Affect.pain>, ‘anxiety’ <Affect.thght>, ‘interruption’ <Affect.int> and ‘expenditure of energy’ <Affect.engy>. On the left, the Transitive forms <Trans> correlate with the use of ‘humour’ <Humour.Hum>, and the Cause and Affects of ‘imposition’ <Affect.imp>, ‘imposing request’ <Affect.Imp_Req>, and ‘tease’ <Affect.tease>.

The results of this correspondence analysis have revealed a clear semantic difference between the transitive verbs and the nominal - gerundive profilings. This is, perhaps, no surprise, but we have captured what kind of lexical semantic differences these two grammatical profilings entail. More importantly, however, is that we can zoom in the level

of granularity. For example, the transitive profilings of this lexical concept can be broken down into a range of different syntactic patterns. Three of these behave in a semantically distinct fashion:

(2) Transitive + Infinitive

- a. Help me hassle them to release the remix.
- b. It's because you keep hassling me to save money.

(3) Transitive + Gerundive

- a. You don't want to be hassled into putting it behind a cut.
- b. My brother finally got hassled into letting me use his laptop after he goes to bed.

(4) Transitive NP for NP (*ask* syntax)

- a. I do not see how this is possible, but this guy keeps on hassling me for sex ...
- b. Examine stunning views of shopping trolleys wedged in the river as the local meths drinkers hassle you for cider money along the river wall.

We can add these more fine-grained distinctions to the correspondence analysis. The three syntactic patterns described above are grouped together and labelled Resultative. This is done to improve legibility upon the plot.

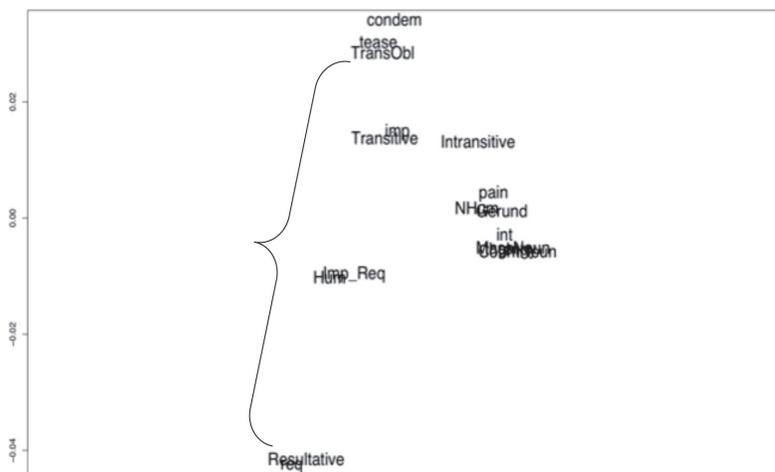


Fig. 2. Interaction of Syntax, Cause, Affect, and Humour
Multiple Correspondence Analysis (Euclidean distance Matrix)

In Figure 2, we see how the general dispersion is maintained, but a cline across the different transitive constructions becomes visible. At one end of the cline, the ‘Transitive Oblique’ <TransObl> correlates with the semantically similar Causes of ‘tease’ <tease> and ‘condemn’ <condem>. Then, the simple ‘Transitive’ <Transitive> and the complex transitive group named ‘Resultative’ <Resultative> are stretched apart. The simple Transitive is associated with the Cause of ‘imposition’ <imp>, where the Resultative transitive constructions correlate with ‘requests’ <req>. Falling exactly between the two data points, and therefore equally associated, we see ‘imposition-request’ <Imp_Req>, a Cause feature where both Cause types are simultaneously involved. Thus, we see how the semantic difference between the two syntax-lexeme pairings are related, forming a continuum from pure and

simple requests through to imposition upon the Patient. Humour <Hum>, a feature shown to be associated with Transitive more generally in Figure 1, remains in-between the two transitive types, showing it is equally associated with all the syntactic sub-types.

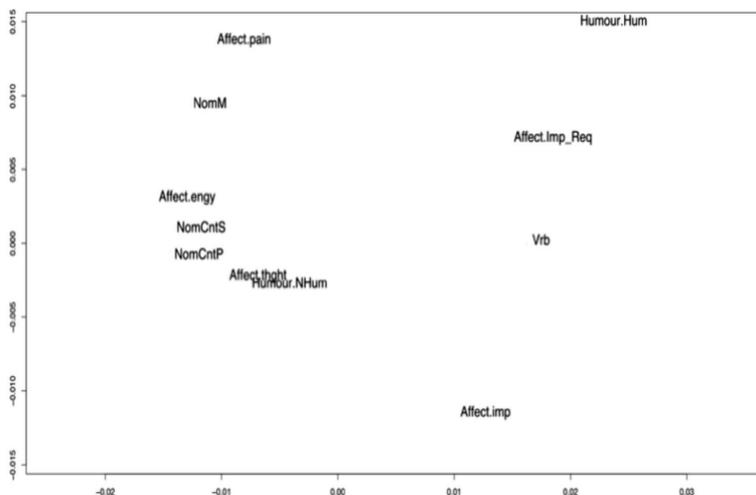


Figure 3. Interaction of Syntax, Cause, Affect, and Humour
Multiple Correspondence Analysis (Euclidean distance Matrix)

Just as we can add more syntactic granularity to the verbal constructions, we can also zoom in on the nominal forms. Figure 3, below, presents the results of a further correspondence analysis where the verbal syntactic variation is conflated and the different nominal profilings are unpacked. In this analysis, we see how the Count Singular <NomCntS> and Count Plural <NomCntP> Nouns cluster together around the feature of ‘non-humour’ (Humour.NHum) and ‘expenditure of energy’ <Affect-engy>. They are distinct from the Mass Noun <NomM> profiling above it, which is distinctly associated with the feature of ‘emotional pain’ <Affect.pain>. These two nominal profilings correlate with clear and different lexical semantics.

Examples (5) and (6) represent the kind of semantic difference Figure 3 captures.

(5) Mass Noun and Affect <emotional pain>

a. I really do feel so shit, I need a clear head, I don't know what I need but I know it isn't this, hassle, pressure, bullying and violence, anger and rage, you can't, no, you can't treat me like this, I know I was never and angel, but there's no need to take me to hell.

b. I would have avoided great hassle and embarrassment had I spotted him earlier. At least now he has got the message. Oh yes, he ruined my life, had his way with me and now he has the cheek to be doing well for himself.

(6) Count Noun and Cause <energy>

a. I rarely paint my nails(It can be such a hassle!)

b. Graphics card is DEAD. I'm 100% sure of it this time though so it shouldn't be too much of a hassle to replace it.

3.2 Statistical confirmation of results

Let us now turn to a confirmatory statistical technique. The results above capture patterns in the data, but they do not indicate whether the patterns are merely chance or how

much of the usage they actually represent. Perhaps these differences in usage exist, but only represent a relatively small number of usage-events. Logistic regression analysis is a confirmatory technique. It models the data and attempts to predict a binary outcome based on the usage-feature analysis. If the model offers an accurate prediction, then we can assume that the analysis has accurately captured the variation in the data. Let us apply this technique to the semantic distinction between the verbal and nominal profilings.

Logistic Regression – Verbal versus Nominal Profiling

Coefficient	S.E.	Wald	Z	P	
Affect=anxiety -5.3006	1.1847	-4.47	0.0000	***	
Affect=emotional pain -3.8490	0.9073	-4.24	0.0000	***	
Affect=energy -6.7874	0.8038	-8.44	0.0000	***	
Cause=condemn		-0.9064	0.9176	-0.99	0.3233
Cause=imposition		-1.0769	0.7303	-1.47	0.1403
Cause=interruption -4.7592		0.9220	-5.16	0.0000	***
Cause=request		2.0800	2.0247	1.03	0.3043
Nominals: 356	Overfit Penalty: 0.05(pentrace)				
Verbals: 266	No multicollinearity				
C: 0.955	2 influential observations removed				
Pseudo R2: 0.81					

The column on the left is a list of the usage-features that were put into the model. The column on the far right is the statistical significance of each feature in predicting the outcome, nominal versus verbal. For these figures, the closer to zero the ‘*p*-value’ (P), the more likely it is that a repeat analysis would achieve the same or better results. Four features prove statistically significant; they are in boldface and have three asterisks to make it easier to identify them. Having established which features are significantly different between the usages of the two forms, we move to the second column – the Standard Estimate (S.E.). This gives the relative strength of the feature in predicting nominal or verbal – the greater the figure, the greater the predictive strength of the feature. All four features predict a nominal form. We see that ‘anxiety’ and ‘expenditure of energy’ are the two more important predictors. We know from the correspondence analysis that ‘energy’ is associated with the count nouns and ‘anxiety’ with the mass nouns. We now know that this correlation is statistically significant and not merely the result of chance. This is especially important with small datasets such as this one. Finally, the overall model is statistically significant and the Affect and Cause Factors successfully enable us to predict the response variable (C – 0.081). The C-value at the bottom of the table can be interpreted as something similar to a percentage, meaning that we can accurately predict to approximately 80% of the time a nominal or verbal example based solely on the Cause and Affect features. In other words, we have captured the semantic difference and shown that this is a robust repeatable finding.

3.3 Summary. Mapping lexico-syntactic semantics

The exploratory technique, correspondence analysis, helped us identify certain usage patterns, specifically the interaction between various syntactic profilings and lexical semantics. We firstly saw that nominal and verbal uses of the lexeme differed considerably in terms of the causes of the events and affects experienced by the patient in those events. These kinds of semantic features are a way of operationalising lexical senses – the clustering of similar causes and similar affects could be argued to represent what were traditionally termed the ‘meanings’ of the word. Seen in this light, we identified different meanings for the different parts of speech. However, the method allows us to zoom in and examine

the relations of the different ‘meanings’ relative to different more fine-grained syntactic contexts, arguably capturing the interaction of the grammatical and lexical semantics. Here, we saw that the use (meaning) of simple transitives and a set of three complex transitive constructions lie on a semantic continuum from imposition through to request and that these uses were often humorous. For the nominal profiling of the concept, we saw two distinct uses (meanings). Mass nouns are associated with more emotionally consequential affects where count noun profiling is more typically associated with light emotional inconsequential affect, caused by the undesired expenditure of energy.

A confirmatory analysis, using logistic regression, was applied to the distinction between the nominal and verbal profiles. The semantic difference between the uses of the two forms was clearly confirmed and the semantic features most important to the distinction identified. The expenditure of energy as a cause and the experience of anxiety were found to be distinctly and significantly associated with the nominal profiling of the concept. We know from the correspondence analysis, that one of the features is associated with the count noun and the other with the mass noun. The combination of these two statistical techniques has helped obtain a picture of the semantic structure of the word *hassle* relative to various morpho-syntactic contexts. The addition of other contexts, such as register or regional variation, is straightforward, but beyond the purview of this study.

4. Conclusions

This short ‘sample’ study on the polysemy of *hassle* has shown how different meanings, associated with a lexeme, exist relative to the lexeme’s syntactic profiling. This interaction of lexical and syntactic semantics needs to be accounted for in linguistic description if we are to accurately capture language use. We saw also that meanings, rather than reified discreet ‘senses’ form continua and are a result of the interaction of different dimensions of language. The exploratory statistics have helped identify these patterns and their multidimensional nature. The confirmatory statistics have helped offer a degree of verifiability to the results. Although repeat analysis or confirmation from the comparison of results from different methods is the best means of verification, statistical modelling is a vital tool in empirical research. Lastly, it is hoped that this short presentation explains one of the ways Cognitive Linguistics is beginning to face and overcome the hurdles that the complexity of language presents. Multifactorial analysis of contextualised language use is proving to be one of the most important methods for approaching language structure holistically.

REFERENCES

1. *Arppe A., Gilquin G., Glynn D., Hilpert M., Zeschel A.* Cognitive corpus linguistics: Five points of debate on current theory and methodology // *Corpora*. – № 5. – 2010. – P. 1-27.
2. *Bartmiński J.* The concept of motherland in present-day European languages. – Lublin : IES-W, 1993.
3. *Chomsky N.* Syntactic Structures. – Paris: Mouton, 1957.
4. *Culioli A.* Pour une linguistique de l’énonciation. Opérations et représentations. – Paris : Ophrys, 1990.
5. *Cuyckens H.* Family Resemblance in the Dutch Spatial Prepositions *Door* and *Langs* // *Cognitive Linguistics*. – № 6. – 1995. – P. 183-207.

6. *Divjak D., Gries St. Th.* Ways of trying in Russian: Clustering behavioral profiles // *Corpus Linguistics and Linguistic Theory*. – № 2. – 2006. – P. 23–60.
7. *Divjak D.* Ways of intending: A corpus-based cognitive linguistic approach to near-synonyms in Russian // *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis*, St. Th. Gries & A. Stefanowitsch (eds.). – Berlin : Mouton, 2006. – P. 19–56.
8. *Divjak D.* Structuring the Lexicon: A Clustered Model for Near-synonymy. – Berlin : Mouton, 2010.
9. *Fillmore C.* Frames and the Semantics of Understanding // *Quaderni di Semantica*. – № 6. – 1985. – P. 222-254.
10. *Geeraerts D.* (ed.) Prospects and Problems of Prototype Theory (Special ed. *Linguistics* 27). – Berlin : Mouton, 1989.
11. *Geeraerts D., Grondelaers S., Bakema P.* The Structure of Lexical Variation. Meaning, naming, and context. – Berlin : Mouton, 1994.
12. *Geeraerts D., Grondelaers S., Speelman D.* Convergentie en Divergentie in de Nederlandse Woordenschat. – Amsterdam : Meertens, 1999.
13. *Gilquin G., Gries St. Th.* (eds). *Corpora and Experimental Methods* (Special ed. *Corpus Linguistics and Linguistic Theory* 5). – Berlin : Mouton, 2009.
14. *Gilquin G., Gries St. Th.* Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory*. – № 5. – 2009. – P. 1-26.
15. *Gilquin G.* Corpus, Cognition and Causative Constructions. – Amsterdam : Benjamins, 2010.
16. *Givón T.* Functionalism and Grammar. – Amsterdam : John Benjamins, 1995.
17. *Glynn D., Fischer K.* (eds). Quantitative Cognitive Semantics. Corpus-driven approaches. – Berlin : Mouton, 2010.
18. *Glynn D., Robinson J.* (eds). Polysemy and Synonymy // *Corpus methods and applications in Cognitive Linguistics*. – Amsterdam : Benjamins, 2010. (In press.)
19. *Glynn D.* Polysemy, syntax, and variation. A usage-based method for Cognitive Semantics // *New Directions in Cognitive Linguistics*, V. Evans & S. Pourcel (eds). – Amsterdam : Benjamins, 2009. – P. 77-106.
20. *Glynn D.* Corpus-driven Cognitive Semantics // *Quantitative Cognitive Semantics. Corpus-driven approaches* / D. Glynn & K. Fischer (eds). – Berlin : Mouton, 2010a. – P. 1-42.
21. *Glynn D.* Synonymy, lexical fields, and grammatical constructions. A study in usage-based Cognitive Semantics // *Cognitive Foundations of Linguistic Usage-Patterns* / H. J. Schmid, S. Handl (eds). – Berlin : Mouton, 2010b. – P. 89-118.
22. *Glynn D.* Testing the hypothesis. Objectivity and verification in usage-based Cognitive Semantics // *Quantitative Cognitive Semantics. Corpus-driven approaches* / D. Glynn, K. Fischer (eds). – Berlin : Mouton, 2010c. – P. 239-270.
23. *Glynn D.* Semantics of Sociolinguistic Variation. A quantitative study of dialect effects on polysemy // *Annual Review of Cognitive Linguistics* (In press a.).
24. *Glynn D.* Polysemy and Synonymy. Corpus method and cognitive theory // *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics* / D. Glynn, J. Robinson (eds). – Amsterdam : Benjamins. (In press b.).

25. *Glynn D.* Mapping Meaning. Usage-based methodology for Cognitive Semantics. – Oxford : OUP. (Forthc. a.)
26. *Glynn D.* Usage-Based Cognitive Models. A corpus-driven study in LIBERTY in 20C America // *Cognitive Linguistics* (Forthc. b.).
27. *Gries St. Th.* Multifactorial Analysis in Corpus Linguistics: A study of particle placement. – London : Continuum, 2003.
28. *Gries St. Th., Divjak D.* (eds). Frequency Effects in Language. – Berlin : Mouton. (Forthc.)
29. *Gries St. Th., Stefanowitsch A.* (eds). Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis. – Berlin : Mouton, 2006.
30. *Gries St. Th.* Corpus-based methods and Cognitive Semantics: The many senses of to run // *Corpora in Cognitive Linguistics. Corpus-based approaches to syntax and lexis / St. Th. Gries & A. Stefanowitsch* (eds). – Berlin: Mouton, 2006. – P. 57–99.
31. *Gries St. Th., Stefanowitsch A.* Extending collocation analysis: a corpus-based perspective on ‘alternations’ // *International Journal of Corpus Linguistics*. – № 9. – 2004. – P. 97-129.
32. *Grondealers S. Speelman D., Geeraerts D.* A case for a cognitive corpus Linguistics // *Methods in Cognitive Linguistics / M. Gonzalez-Marquez, I. Mittleberg, S. Coulson & M. Spivey* (eds). – Amsterdam : Benjamins, 2007. – P. 149-169.
33. *Heylen K.* A quantitative corpus study of German word order variation // *Linguistic Evidence: Empirical, theoretical and computational perspectives / S. Kepsner, M. Reis* (eds.). – Berlin : Mouton, 2005. – P. 241-264.
34. *Hilpert M.* Germanic Future Constructions. A usage-based approach to language change. – Amsterdam : Benjamins, 2008.
35. *Hilpert M.* Collocation Analysis: Measuring associations between constructions and lexical elements // *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics / D. Glynn, J. Robinson* (eds.). – Amsterdam: Benjamins. (In press.).
36. *Janda L., Solovyev V.* What constructional profiles reveal about synonymy: A case study of the Russian words for SADNESS and HAPPINESS // *Cognitive Linguistics*. – № 20. – 2009. – P. 367-393.
37. *Kilgarriff A.* I don't believe in word senses // *Computers and the Humanities*. – № 31. – 1997. – P. 91-113.
38. *Lakoff G.* Women, Fire, and Dangerous Things. What categories reveal about the mind. – London : UCP, 1987.
39. *Lakoff G.* Moral Politics: What Conservatives Know That Liberal Don't. – London : UCP, 1996.
40. *Langacker R.* Foundations of Cognitive Grammar. – Vol. 1. Theoretical prerequisites. – Stanford : SUP, 1987.
41. *Langacker R.* A usage-based model // *Topics in Cognitive Linguistics / B. Rudzka-Ostyn* (ed.). – Amsterdam : Benjamins, 1988. – P. 127-161.
42. *Langacker R.* Grammar and Conceptualization. – Berlin : Mouton, 1999.
43. *Lewandowska-Tomaszczyk B., Dziwirek K.* (eds.). Studies in Cognitive Corpus Linguistics. – Frankfurt/Main : Lang, 2009.

44. *Saussure F. de.* Cours de linguistique générale. – Paris : Payot, 1995 [1916].
45. *Schmid H.-J.* Cottage and Co., Idea, Start vs. Begin. Die Kategorisierung als Grundprinzip einer differenzierten Bedeutungsbeschreibung. – Tübingen : Niemeyer, 1993.
46. *Schmid H.-J.* English Abstract Nouns as Conceptual Shells. From corpus to cognition. – Berlin : Mouton, 2000.
47. *Sinclair J.* Corpus, Concordance, Collocation. – Oxford : OUP, 1991.
48. *Stefanowitsch A., Gries St. Th.* Collocations: Investigating the interaction of words and constructions // International Journal of Corpus Linguistics. – № 8. – 2003. – P. 209-243.
49. *Stefanowitsch A., Gries St. Th.* (eds). Corpus-Based Approaches to Metaphor and Metonymy. – Berlin : Mouton, 2006.
50. *Szmrecsanyi B.* Morphosyntactic Persistence in Spoken English. A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis. – Berlin : Mouton, 2006.
51. *Taylor J.* Linguistic Categorization: Prototypes in linguistic theory. – Oxford: Clarendon Press, 1989.
52. *Tummers J., Heylen K., Geeraerts D.* Usage-based approaches in Cognitive Linguistics: A technical state of the art // Corpus Linguistics and Linguistic Theory. – 2005. – № 1. – P. 225-261.
53. *Wierzbicka A.* Lexicography and Conceptual Analysis. – Ann Arbor : Karoma, 1985.
54. *Wulff S.* Rethinking Idiomaticity. A usage-based approach. – London : Continuum, 2009.
55. *Wulff S., Stefanowitsch A., Gries S. Th.* Brutal Brits and persuasive Americans: Variety-specific meaning construction in the into-causative // Aspects of Meaning Construction / G. Radden, K.-M. Köpcke, T. Berg, P. Siemund (eds). – Amsterdam : Benjamins, 2007. – P. 265-281.
56. *Zeschel A.* (ed.). Usage-Based Approaches to Language Processing and Representation (Special ed. Cognitive Linguistics 19). – Berlin : Mouton, 2008.
57. *Воркачев С. Г.* Счастье как лингвокультурный концепт. – Москва : Гнозис, 2004.
58. *Кубрякова Е. С., Янко, Т. Е.* (ред.). Язык и культура: Факты и ценности: К 70-летию Юрия Сергеевича Степанова. – Москва : Языки славянской культуры, 2001.
59. *Степанов Ю. С.* Константы. Словарь русской культуры : Опыт исследования. – Москва : Школа «Языки русской культуры», 1997.

Стаття надійшла до редколегії 11.04.2011

Принята до друку 15.04.2011

**CORPUS-DRIVEN COGNITIVE LINGUISTICS.
A CASE STUDY IN POLYSEMY****Дилан ГЛИН***Senior research fellow
Centre for Language and Literature
Lund University, Sweden EU*

Стаття присвячена квантитативним корпусним методам когнітивної лінгвістики, в якій використовується модель мови, заснована на її використанні. Ця модель дозволяє здійснювати концептуальний аналіз на матеріалі корпусних даних. Якщо припустити, що граматику є результатом повторного вживання (закріпленості) мовних одиниць, то граматику (концептуальну структуру) мови можна «зобразити» на основі аналізу повторюваних комбінацій ознак, що характеризують вживання цих одиниць у великій кількості контекстів. Один із методів такого аналізу застосовується в цьому дослідженні для вивчення взаємодії морфо-синтаксичної та лексичної семантики. У рамках даного методу використовуються два статистичних прийоми, за допомогою яких здійснюється пошук і підтвердження повторюваних комбінацій ознак мовного вживання лексеми *hassle* в англійській мові.

Ключові слова: когнітивна лінгвістика, квантитативні корпусні методи, концептуальний аналіз.

**CORPUS-DRIVEN COGNITIVE LINGUISTICS.
A CASE STUDY IN POLYSEMY****Дилан ГЛИНН***Senior research fellow
Centre for Language and Literature
Lund University, Sweden EU*

Данная статья посвящена квантитативным корпусным методам когнитивной лингвистики, в которой используется модель языка, основанная на его употреблении. Эта модель позволяет проводить концептуальный анализ на материале корпусных данных. Если предположить, что грамматика является результатом повторного употребления (закрепленности) языковых единиц, то грамматику (концептуальную структуру) языка можно «изобразить» на основе анализа повторяющихся комбинаций признаков, характеризующих употребление этих единиц в большом количестве контекстов. Один из методов такого анализа применяется в настоящем исследовании для изучения взаимодействия морфо-синтаксической и лексической семантики. В рамках данного метода используются два статистических приема, с помощью которых осуществляется поиск и подтверждение повторяющихся комбинаций признаков языкового употребления лексеми *hassle* в английском языке.

Ключевые слова: когнитивная лингвистика, квантитативные корпусные методы, концептуальный анализ.