

The semantics of sociolinguistic variation. A quantitative study of dialect effects on the polysemy of *annoy*

Dylan Glynn
University of Lund

Abstract

According to the usage-based model of language propounded by Cognitive Linguistics, extralinguistic context is integral to language structure. Despite this, it is often sidelined in Cognitive Linguistic research. Polysemy represents a foundational line of research in Cognitive Linguistics, and this study demonstrates why extralinguistic concerns should be integrated into polysemy analysis. Following Gries (2006) and Glynn (2009), the analysis takes a corpus-driven multifactorial approach to semasiological structure. It applies this method to the lexeme *annoy* in British and American English. The data are taken from a non-commercial corpus consisting of personal on-line diaries and are analysed for a range of formal and semantic features. The statistical exploration of the results, using Multiple Correspondence Analysis, identifies three basic senses of the terms as clusters of usage features. One of these three senses is found to be canonical with the other two senses less central to the usage of the lexeme. The two less typical senses appear to be regional variations, one use being typical of the American and the other of British usage.

Keywords: polysemy, sociolinguistics, corpus linguistics, correspondence analysis

1. Introduction. Language context and usage-based polysemy

Variation in social strata, gender, ethnicity, age, and dialect directly influences the use of language.¹ In a usage-based approach to language, this entails that such sociolinguistic factors have an effect on semantic relations such as polysemy, synonymy, or metaphor. We must, therefore, account for this variation in semantic analysis. The application of this theoretical prerequisite to language analysis is tied to methodological advances. Within

2 *The semantics of sociolinguistic variation - draft*

Cognitive Linguistics, the necessity of empirical methods has become widely accepted (Geeraerts 2006; Gibbs 2007; Croft 2008), facilitating the inclusion of sociolinguistic parameters in its approach to language description. This study examines the effect of regional variation on semantic structure and considers empirical techniques for capturing that effect. Using a large non-commercial corpus, the study investigates the effect of British and American dialect variation on the polysemy of the lexeme *annoy*.

The study of polysemy and the development of analytical apparatuses designed to capture patterns of conceptual structure belong to the foundational lines of enquiry within the cognitive paradigm. Work on semantic variation and its categorisation, such as Vandeloise (1985), Lakoff (1987), Geeraerts (1989), and Cuyckens (1993, 1995), is arguably the corner stone of Cognitive Semantics. However, despite the importance and success of these contributions, the conceptual reality of their results has been brought into question (Sandra & Rice 1995). In response to this, Zelinsky-Wibbelt (2000: 145) argues that the problem is a methodological one. Her position is simply that Cognitive Linguistics is trying to model lexical semantic structure without an analytical structure that accounts for the role of context.

Deane (1988, 2005), Vandeloise (1994), and Glynn (2003) have each focused on different elements of this point demonstrating that much semantic variation identified in Lexical Network Analyses (Lakoff 1987) is a result of factors that traditionally lie outside the domain of lexical semantics. This argument is echoed by Evans (2004, 2005), who suggests that we need to reinstate a distinction between the 'denotational' lexical semantic concept and a broader pragmatic-conceptual level of meaning. We will not take this position, but rather assume that the proposal, *a priori*, of a distinction between meaning types is unnecessary. Cognitive Linguistics is a usage-based approach to language (Langacker 1988) which maintains that all knowledge relevant to the correct usage of a linguistic form is part of the meaning of that linguistic form (Lakoff 1987).

The task for us, as empirical scientists, is to develop methods to make reliable generalisations about precisely that usage. We should not seek to simplify the object of study by dividing parameters of use into different types of meaning or levels of linguistic structure. Instead, we can treat them as different factors that an individual, pre-consciously, uses to determine the choice or interpretation of a word. Understanding the role of sociolinguistic effects in these terms permits us to seamlessly introduce their study into both the theoretical model of meaning propounded by Cognitive Lin-

guistics and the methodological apparatus espoused by corpus linguistics. This study attempts precisely this, focusing on one particular sociolinguistic factor, that of regional variation. We show why this is essential in the study of semantic structure, but also we consider some simple techniques for identifying and representing such patterns of language use.

1.1. Multifactorial and quantitative research in Cognitive Linguistics

Although Langacker (1987) stresses that a linguistic unit is form and meaning, just as Lakoff (1987) stresses that meaning is our encyclopaedic knowledge of the world, the full implications of approaching encyclopaedic meaning in the context of usage have not been fully appreciated. Cognitive Linguistics has long argued that we cannot treat linguistic phenomena as museum pieces, as decontextualised artefacts of language. These arguments were directed at the modular theories of language that sought to reduce linguistic complexity by imposing semantic and formal ‘modules’ upon language structure. However, in practice, if not in theory, Cognitive Linguistics continued this tradition by all but completely ignoring the usage-context of language. This is now changing.

One of the developments that permit this change is the recent burgeoning of quantitative corpus-driven methodology (Geeraerts 2005, 2006; Tummers *et al.* 2005; Gries & Stefanowitsch 2006; Stefanowitsch & Gries 2006; Grondelaers *et al.* 2007; Wulff *et al.* 2007; Heylen *et al.* 2008; Zeschel 2008; Hilpert 2008; Wulff 2010; Gilquin 2009, 2010; Glynn 2010a, 2010b; Divjak 2010; Glynn & Fischer 2010; Glynn & Robinson in press).² This pullulation of interest is timely and represents fertile ground for finally bringing socio-functional concerns to the fore of cognitive research. Sociolinguistic factors lend themselves well to quantitative analysis and the sociolinguistic tradition has a rich methodological heritage to lend to cognitive lines of enquiry. This trend within the cognitive paradigm is represented by Geeraerts (2005), Wulff *et al.* (2007), Kristiansen & Dirven (2008), Geeraerts & al. (2010), and Speelman & Geeraerts (2010). However, since it is not always clear how corpus linguistic research can inform Cognitive Semantics, we need to overtly state how usage-based method can be used to answer questions about conceptual structure.

1.2. Indices of semantic structure – a usage-based method

Langacker's (1987) theory of entrenchment is crucial to understanding how corpus-driven methods inform Cognitive Semantics. Langacker (1987: 59-60) stresses that frequency is the motivating force behind entrenchment, that the more often a speaker uses a form-meaning pair, the more automated and 'entrenched' it will be in the speaker's personal knowledge of that language. This is Cognitive Linguistics' theory of grammaticality. Corpus linguistics examines large numbers of expressions in natural language and looks for usage-patterns in that language. It identifies those patterns in terms of relative frequency in the combination of uses and forms. Therefore, according to Langacker's theory of grammaticality, corpus linguistics is mapping the grammar of a language.

Moreover, Lakoff (1987) argues that linguistic structures are indices of conceptual structures, where conceptual structures are understood as learnt relations between experienced phenomena. Most Cognitive Linguistic research, regardless of methodology, assumes this premise and, indeed, it is fundamental to any semantics-driven model of language. In accepting this hypothesis, we can extend the premise to the study of corpora with far-reaching implications.

If linguistic structure can be ascertained through the identification of relative frequency, and if linguistic structures are indicative of underlying semantic structures, then corpus techniques can be used to indirectly map the conceptual structure of a language and its speech community. Some Cognitive Linguistic scholars, such as Schmid (2000: 39) and Gries (2003), have gone further and argued that we can make claims about cognition based on corpus analysis. Their position, the Corpus to Cognition Principle, has met important criticism (Schmid 2010). However, a Corpus to Concept Principle is much more tenable.

Principle of Corpus to Concept

Patterns of language use index semantic structures of a language, which in turn, reflect conceptual structure.

This principle represents a basic link between corpus analysis and the research of Cognitive Semantics. Moreover, since corpora represent an excellent means of capturing sociolinguistic structure, it permits us to incorporate context of use into our analyses. This means we do not only have a

picture of conceptual structure, but of conceptual structure relative to context and function – a Socio-Cognitive Semantics.

2. *Annoy*. Semantic variation relative to regional variation

Following Gries (2006) and Glynn (2009), this study takes a multifactorial approach to polysemy. It advances upon the previous work by examining the role of sociolinguistic factors in semantic variation. Specifically, we consider the effects of regional variation. The lexeme *annoy* is chosen since it represents the kind of word where one would expect little sociolinguistic variation in use. The item, of Latinate origins, does not belong to the colloquial registers where typically regional variation is most pronounced, nor does it belong to formal registers of language where text-type might influence the uses of the word. In simple terms, one would expect the item to be neutral with regards to possible sociolinguistic effects, such as dialect, gender, register and so forth. In this respect, *annoy* serves as a good test-case for the importance of such phenomena.

The *Oxford English Dictionary* defines the usage thus:

1. Verb intrans: Be hateful, be cause of trouble. Archaic only
2. Verb trans: Cause slight anger or mental stress
3. Molest, injure, harass
4. Damage something material. Dialectal only

Definitions between the British *Oxford English Dictionary* and the American *Webster's Dictionary* tend to differ substantially. The differences between the two in this case are not remarkable. Below is the definition offered by the American *Webster's Dictionary*:

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

Let us consider how these definitions compare to the semantic range identified using corpus-driven multivariate techniques on data distinguished for British and American usage.

2.1. Corpus and annotation

The corpus was compiled by D. Speelman at the research unit of Quantitative Lexicology and Variational Linguistics, the University of Leuven. It is a large unparsed corpus made up of on-line personal diaries, taken from the popular ‘blog’ service entitled LiveJournal. This service has the important characteristic that authors can indicate not only their place of origin, but also their secondary school and its address. This gives one a reasonably sure means of establishing the regional variation of the data. The corpus is thus stratified for dialect: one-third British and two-thirds American usage.

The language of the on-line diaries is familiar in register and takes a semi-narrative form. Unlike traditional diaries, these diaries are designed to be read by many people, indeed potentially millions. Moreover, readers can, and do, respond to what is written. This leads to quite a unique style that differs markedly from the traditional diary genre. Such a style could certainly be a source of variation in use. For this reason, the examples are annotated for stylistic concerns such as true narrative use, which occurs when amateur creative writers use the service as a forum to discuss their writing. Also, occasionally, there are instances of “cutting and pasting” from news press articles and so on. In order to control for such variation, such occurrences are not included in the sample.

Other social dimensions, such as gender, age, and socio-economic class are impossible to determine with any degree of certainty. Although, in readings the diaries, it seems that if there is any gender bias, it is towards women. Social class is harder to determine. However, one clue to this lies in the discursive theme, or topic of discourse, which is annotated. The themes are typically concerned with family, friends, university, work, computers, and the on-line ‘blogging’ community itself. If we make the assumption that younger and financially comfortable individuals are more likely to be attending tertiary education, then we can deduce that the majority of users belong to the younger middle classes. There are also many references to secondary schooling and parents, which indicates with some certainty a large number of adolescent users. However, with the recent democratisation of the Internet in Anglo-Saxon society, it would be reckless to suppose that these adolescent users belong to the lower, middle, or upper classes.

The study is based on a sample of 500 occurrences, divided equally between American and British instances. Although this analysis will only consider a few variables, the occurrences were manually annotated for a

wide range of formal, semantic, and sociolinguistic factors; in total some 20 variables across 120 features. One of the challenges that a quantitative approach to polysemy faces is the abstract nature of most linguistic concepts. One solution that is commonly employed in corpus linguistics is to examine formal variation as an indirect index for variation in use. However, for an item such as *annoy* there is little informative morpho-syntactic variation. Therefore, following the Cognitive Semantic work of Dirven & al (1982), Verschueren (1985), and Rudzka-Ostyn (1989, 1995), we focus on features of the event-frame as a means for operationalising the semantic feature analysis. In this approach, features such as the ‘animacy’, ‘abstractness’, or ‘familiarity’ of the Agent and Patient are used as an indirect method for identifying semantic variation. This is compared with other direct and indirect factors, such as the Cause of the event, the Affect for the Patient as well as the Thematic Topic of Discourse. The factors and their usage features are discussed as we consider their effects on, and interaction with, the variation in the following sections.

2.2. Multifactorial Analysis. Usage features as indices of semasiological structure

In order to capture the interaction between the various usage factors and the regional variation of dialect, a wide range of statistical tools are available. This study employs Multiple Correspondence Analysis, combined with two simple techniques for the visualisation of contingency tables. Correspondence Analysis computes distances as a way of representing correlations, or correspondences, within data. To do this, it uses a Euclidean or a chi-square metric, taking the frequency of feature co-occurrence and converting this to distance, which can then be visualised on a two-dimensional plot. Glynn (in press) explains the technique in more detail. In the following analyses, the Euclidean metric is employed. In interpreting the plots that it produces, one must remember that the visualisations are not drawn but are generated by the mathematical algorithm. Moreover, the visualisations conflate a multidimensional space to two dimensions. This can lead to distortions and interpretations must be made with caution. On the plots, distance, although representative of correlation, is not absolute but relative.

2.2.1. *Actor types*

The first factors that we will consider are the Actor types of the event-frame. The first Correspondence Analysis, Figure 1, visualises the correlations between Agent type, Patient type, and Dialect. The Actor types annotated and presented here in the analysis include Human Agents that are known and specified in the event <Ag.HumSp> and also Human Agents that are unknown to the speakers <Ag.HumNSp>. The two equivalent Patient types are identified <Pat.HumSp>, <Pat.HumNSp>. Three Inanimate Agent types are also included: ‘things’ <Ag.Thg>; ‘events, activities, or occurrences’ <Ag.Evnt>; and ‘abstract states of affairs’ <Ag.AbSoA>. The two dialect variants are indicated by <UK> and <US>.

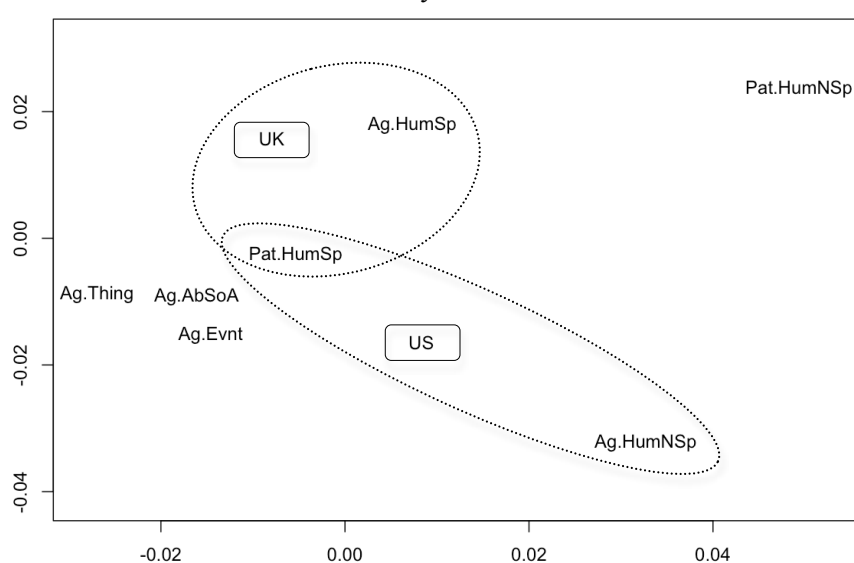


Figure 1. Multiple Correspondence Analysis of *annoy*. Agent, Patient, and Dialect

Firstly, notice that the British <UK> and American <US> data points are quite central to the plot. Placed squarely between them is the Patient type ‘specified human’ <Pat.HumSp> representing the canonical Patient type of ‘annoy events’. In the data set, it was typically the speaker and object of the event, or *me*. This feature is so basic to the use of *annoy* in both dialects that it lies neatly between the two.

The clustering of all three non-animate Agent types is intuitively sound due to their semantic similarity in contrast to animate agents. This grouping of the inanimate Agent types and the ‘specified human’ Patient should be

seen as the canonical event scenario. The ‘non-specified human’ Patient <Pat.HumNSp> is of such a low occurrence that it shows no clear associations.

Most interestingly, for our concerns, the plot shows some regional variation in usage. The marked disassociation of ‘specified human’ Agents <Ag.HumSp> and ‘non-specified human’ Agents <Ag.HumNSp> appears to be a result of a difference between the dialects. The plot shows a high degree of correlation between ‘specified human’ Agents and the British use of *annoy*, relative to the American usage, where the reverse is true and ‘non-specified human’ Agents <Ag.HumNSp> are associated with the American data point. The distinction is not statistically significant, possibly because of the relatively small number of ‘non-specified human’ Agents in the data combined with the fact that ‘specified human’ Agents, although much more typical of the British usage, are not distinctly typical. That is to say, ‘specified human’ Agents also often occur in the American data set. This is further investigated below.

Remembering that the working hypothesis is that characteristics of usage, such as Agent and Patient type, are an indirect index of semantic structure, we may posit that the kind of usage patterns revealed in Correspondence Analysis captures variation in semantic structure such as polysamy. Let us now add more variables in the hope of producing a more complex picture of the usage of *annoy* across the two dialects.

Figure 2 visualises the results of the previous Correspondence Analysis with two added variables, Agent Person and Patient Person. These variables are the grammatical number of the two actors. Due to data sparseness, Patient Number is reduced to ‘first person’ <PatPers.P1st> and ‘non-first person’ <PatPers.NP1st>, second and third person being conflated. For the same reason, no distinction is made between plural and singular.

In this plot, again, we see the two dialect points lying in the centre with a range of canonical features between them. Immediately, the association between the different inanimate Agents on the right, ‘things’ <SubTyp.Thg>, ‘abstract state of affairs’ <SubTyp.AbSoA>, and ‘events - activities’ <SubTyp.Evnt>, is clear and intuitively sound (i). Although difficult to read, these overlap with the ‘specified human’ Patient <Pat.HumSp>, the ‘first person patient’ <PatPers.P1st>, and ‘third person’ Agent <AgPrs.A3rd>. This collection of features surely represents what is most typical in the data for both dialects. Since inanimate Agents are almost exclusively ‘third person’ Agents, this clustering is to an important degree a result of redundancy between the factors. Nevertheless, the association of these Agent types with the ‘first person’ Patient captures the ca-

nonical ANNOY event. This cluster represents a scenario where some ‘thing’ or some ‘event’ affects the speaker and the speaker describes this ‘affect’ with the verb *annoy*.

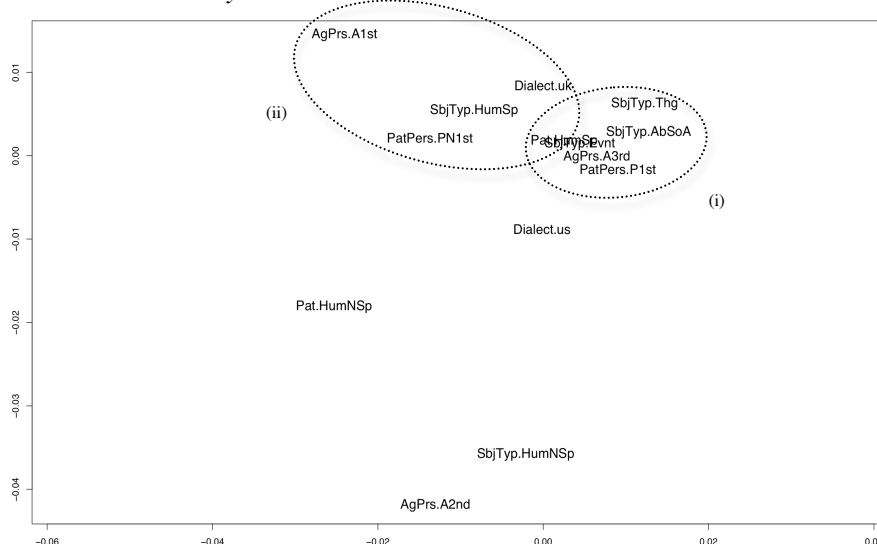


Figure 2. Multiple Correspondence Analysis of *annoy*.
Agent Person, Patient Person, Subject Type, Patient Type, and Dialect

Examples (1a) - (1d) are representative of this cluster of features.

- (1) a. Wow it's finally working again so i figured i best update quick before either livejournal or my pc decides to stop working and really *annoy* me.
- b. We're back dealing with things going wrong, things being delayed, and all the rest of the day to day things that *annoy* the crap out of me.
- c. I started off the week in the same frame of mind that I was in last week - depressed, the slightest thing *annoying* me, crying over nothing, etc.
- d. Currently straightening my hair for photo day tomorrow.. but its not going how I want it to.. its going poofy and *annoying* me!

A word of warning must be introduced at this stage. Redundancy between factors results in serious problems for some statistical analyses. This redundancy, or multicollinearity in statistical terms, cannot be included when modelling data for hypothesis testing. However, in exploratory analysis,

such as Correspondence Analysis, it does not affect the results, only their interpretation. Nevertheless, one must take every caution to consider this point when interpreting the plots.

An interesting line of investigation and one that is less intuitively obvious is the correlation between ‘specified human’ Agents <SubTyp.HumSp> and ‘non-first person’ Patients <PatPers.PN1st> indicated by (ii) on the plot. This pairing of features lies in-between the two dialects suggesting it is common to both varieties. Two examples of this pairing include:

- (2) a. The simplest of things managed to *annoy* him and he was more inebriated than the birthday boy.
- b. I try to talk her out of it whenever she mentions it. It probably *annoys* her because she quits talking to me after I do it

Although it is possible that these examples represent a specific usage type, it is also possible that these data points are caught in-between the canonical group (i) and the ‘first person’ Agent <AgPrs.A1st>, which is distinctly associated with the British data point <Dialect.uk>. This is the greatest difficulty in interpreting correspondence plots. If given features are associated with other features that are dispersed on the plot, then those features lie in-between, giving an impression that they are not associated. Only careful interpretation can ward off false deductions at this point. In this situation, such a ‘stretching’ between different associations is likely the cause of the distinction in the plot. There is necessary redundancy between ‘specified human’ Agents <SubTyp.HumSp> and ‘first person’ Agents <AgPrs.A1st>. This multicollinearity, that ‘specified human’ Agents <SubTyp.HumSp> necessarily coincide with the distinctly British ‘first person’ Agents <AgPrs.A1st>, may explain why the data point ‘specified human’ Agents lies in-between the two regions of the plot, where, in reality, it may be part of the canonical usage represented by cluster (i). This latter explanation seems likely, but closer investigation is needed to be sure.

For our purposes, the association between the British <Dialect.uk> and ‘first person’ Agent <AgPrs.A1st> data points is of more concern. This association may explain the other association between ‘specified human’ Agents and British usage seen in Figure 1. Examples (3a) – (3b) are typical of this usage.

- (3) a. So gotta go *annoy* Steven by sending him a message while he is more than likely trying to stop a boy from biting him lol
- b. I'm hyper and feel like being impish and silly with people but instead i'm stuck here with no-one to *annoy* and no-one to be an-

noying with! Hiss. I want to build a water-slide down the stairs
and roll people around in...

These kinds of examples are typical of the British data, not the American, and are partial indicators of the different pattern of usage that we will see emerging across the two regional varieties.

The plot reveals two other interesting correlations. Firstly, both ‘non-specified human’ Agents <SubTyp.HumNSp> and ‘non-specified human’ Patients <Pat.HumNSp> appear to be distinctly typical of the American usage. Although they do not cluster with the American data point, they are plotted a long way from the British point relative to the American point. We will consider this more closely below, but it is related to an important thematic difference in the British and American usage. The American use of *annoy* is more often associated with discussion about society and less with personal topics. It must be remembered that the examples come from on-line personal diaries and so discussion about society at large is quite marked, the vast majority of the discussion being about intimate subjects. Secondly, the ‘second person’ Agent <AgPrs.A2nd> data point is as far from the British English data point as possible, relative to the position of the American data point. This demonstrates that *you annoy* is markedly un-British in the sample. We also investigate this below.

Other than Correspondence Analysis, another way of representing simple correlations is with a mosaic plot. Let us take a second look at Actor number and Patient number relative to Dialect. Figure 3 is an example of this kind of ‘mosaic’ representation. Although mosaic plots can represent multiway contingency tables, they become difficult to interpret if one adds more than three or four variables. In these plots, ‘block size’ represents degree of association relative to the other associations, such that larger blocks in the mosaic indicate more important correlations.

Figure 3 clarifies the difference we saw between the dialects in the Correspondence Analysis plotted in Figure 2. There, we saw that the grammatical person and type of Agents and Patients differed relative to the dialects; this plot captures the relationship between specifically grammatical person and dialect. The plot is divided into two columns and three rows of unequal size, proportional to the relative frequency of the given features, but this, in turn, is divided into British and American frequencies. So, the proportion of the different Agent and Patient persons is depicted relative to the two dialects. In the left column, we see the large spaces that represent the most common ‘third person’ Agent <A3rd> combined with the ‘first person’ Patient <P1st>, relative to the two dialects. These represent the examples where ‘someone’ or ‘something’ external to the utterance is ‘an-

noying’ the speaker. This is the canonical event type and is evenly distributed across the two dialects; its central status corresponds with what we saw in the Correspondence Analysis above. The fact that there is no difference between the dialects at this point, adds weight to the argument that the ‘specified human’ Agent <SbjTyp.HumSp>, which was ambiguously placed in plot of Figure 2, does in fact belong to the central cluster representing the canonical usage.

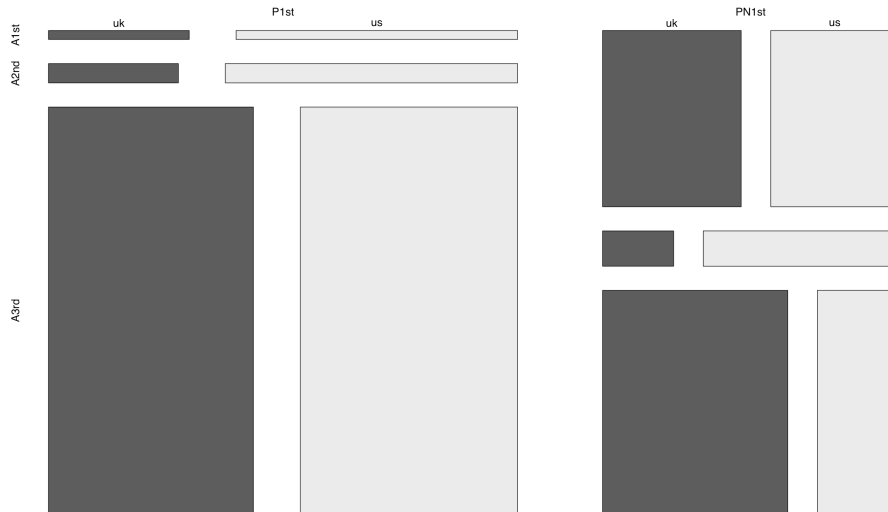


Figure 3. Mosaic Plot *annoy*. Agent person, Patient person, and Dialect

In Figure 3, for the ‘first person’ Patient <P1st> (above the ‘third person’ Agent <A3rd>), we see the dialect distribution of the less frequent ‘first person’ Agent <A1st> and ‘second person’ Agent <A2nd> combined with the ‘first person’ Patient <P1st>. The combination of ‘first person’ Agent and ‘first person’ Patient is a result of a small number of reflexive uses. The infrequency of such examples means that we cannot make generalisations about the use of this form in this data set. However, for the ‘second person’ Agent, it is useful to compare the British and American usage, indicated by the left and right columns. For both the ‘first person’ Patient <P1st> and the ‘non-first person’ Patient <PN1st>, the use of the ‘second person’ Agent (*you*) <A2nd> is markedly American. This parallels with what we saw in the Correspondence Analysis, but is more clearly captured by the mosaic plot. However, notice that in the mosaic plot, the ‘first person’ Agent <A1st> combined with ‘non-first person’ Patient <PN1st>, in the second column, is only marginally more typical of the British usage. In the plot of the Correspondence Analysis, the ‘first person’ Agent <AgPers.A1st>

seems to be distinctly non-American in usage. This relative contradiction could be a result of various factors in the Multiple Correspondence Analysis, including the tug of other variables included or simple data sparseness. We must be aware that these types of exploratory analysis are designed for generating hypotheses and the results still need confirmation through statistical modelling, through corroboration by comparing the results with those gleaned from other methods, or through repeat corpus analysis on different datasets.

Another useful way of representing correlations is via an association plot. It has already been employed for the study of semantic structure in Cognitive Linguistics (Gries & David 2007). Figure 4 represents the Agent type relative to the two dialects. The important difference between American and British in the use of ‘non-specified human’ Agents <HumNSp> is clearly depicted and is statistically significant ($p < 2.2e-16$). This verifies what we saw in Figure 2, where the Correspondence Analysis placed the ‘non-specified human’ Agent <SubTyp.HumNSp> data point as distinctly American relative to the British data point. This kind of verification is important since, hypothetically, the data point could have been drawn to another unrelated point, as in this case above with the ‘first person’ Agent <AgPrs.A1st>.

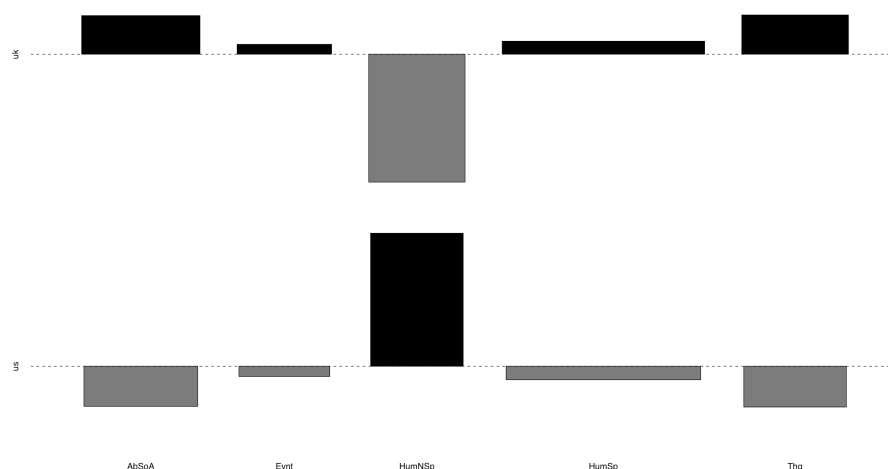


Figure 4. Association plot *annoy*. Agent type and Dialect

The mosaic plot in Figure 3 and the association plot in Figure 4 have helped us verify what we saw in the Correspondence Analysis, that the

‘second person’ Agent and the ‘non-specified human’ Agent are two Actors that are distinctly typical of the American usage of *annoy* in this data set.

Let us now consider one last description of the Actor type and Dialect associations in the form of a more complex Mosaic plot. In Figure 5, the five Agent types are depicted on the horizontal axis relative to Patient on the vertical axis with Dialect distributed between them. The canonical associations of ‘specified human’ Agent <HumSp> and ‘specified human’ Patient are clear in the third column, just as their equal distribution across the two dialects is clear.

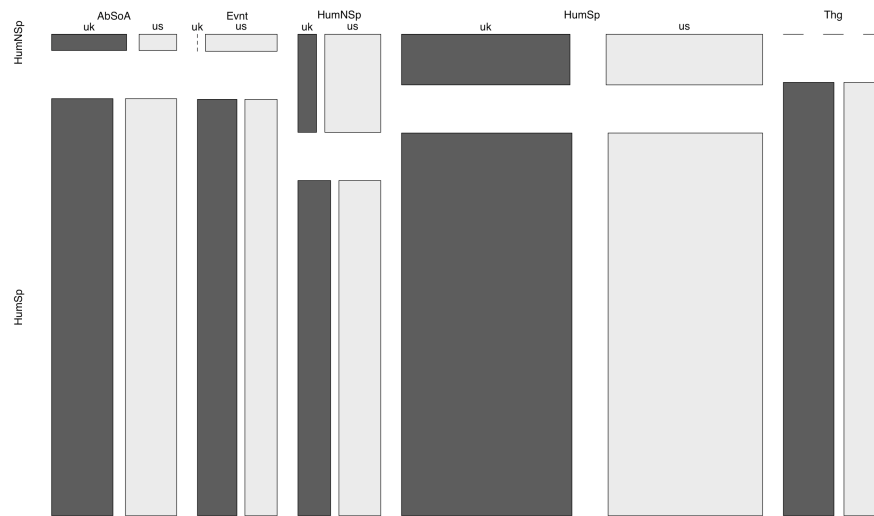


Figure 5. Mosaic plot *annoy*. Agent type, Patient type, and Dialect

Next, we see again the association of the ‘non-specified human’ Agent <HumNSp> and the American usage <us>. Note how this kind of Agent co-occurs relatively frequently with ‘non-specified human’ Patients <HumNSp>. This correlation between Actor types is typical of the American data. Example (4) represents this kind of occurrence.

- (4) a. A new federal law states that when you *annoy* someone on the Internet, you must disclose your identity.
- b. guys that like you who constantly *annoy* you thinking it might work.

Notice as well that the Agent type ‘thing’ <Thg> does not occur with ‘non-specified human’ Patients <HumNSp> and that there are no examples of British ‘event-activity’ Agents <Evtnt> affecting ‘non-specified human’

Patients <HumNSp>. This, of course, could just be due to the limited size of the data set.

What do these differences between the dialects tell us? Recall our hypothesis that Actor types can serve as an indirect method for capturing semantic structure. We can, therefore, attempt to interpret these differences in terms of semasiological variation. It seems intuitively sound that a known third person affecting the subject is the canonical event type for the semantic frame and that, in turn, this is shared between the dialects. However, the differences in less canonical event types may be indicative of differences in usage, or conceptual structure, between the dialects. Let us follow this line of argumentation.

The mosaic plots have shown two clear tendencies in the data. Firstly, the lexeme *annoy* is used commonly with Agents that are ‘non-specified humans’ in American English, where in British English the ‘annoyers’ tend to be specified individuals. Secondly, when the Agents are specified, another relatively clear difference emerges. In British, the first person speaker often coincides with the Agent, where in American, this is rare. Moreover, in American, we often find the Agent is the second person. Do these findings indicate semantic structure, and therefore, difference between the semasiological structures of the two dialects? By looking at some examples, we will be able to see the type of semantics these indices of usage represent. We can summarise these differences and consider some examples of each in order to appreciate the difference in usage. Firstly, examples (5a) - (5b) are typical of events where the ‘annoyer’ and ‘annoyed’ are both unspecified humans. These examples were typical of the American data and not the British.

- (5) a. Whoever...utilizes any device or software... without disclosing his identity and with intent to *annoy*, abuse, threaten, or harass any person...shall be fined under title 18 or imprisoned not more than two years, or both.
- b. Call Your Congressional Rep! *Annoy* your Senator!

Secondly, and also typical of American, the ‘annoyer’ is the second person and the ‘annoyed’ is the first person:

- (6) a. Please stop thinking that everything i write is about you. that i live and breathe each day to hate you. I DONT HATE YOU...but you *annoy* the shit out of me.
- b. "Harry's right, Risa, lay off Daisuke for a while. You're starting to *annoy* me as well," Satoshi agreed quietly.

Thirdly, typical of the British usage are examples where the ‘annoyer’ and the speaker are the same person:

- (7) a. Basically, to make a sale, i *annoy* the fuck out of people! like, majorly. eventually they get pissed off, and hang up, but you get the dumb fucks who agree to buy
- b. Eh, I'm off to *annoy* my sexy doggie-woggie with hugs and kisses. *insert evil laugh.*
- c. James from Riversdales joined me of my unofficial quest to *annoy* Jim as much as possible.
- d. So I think all that's left is to get drunk and eat turkey and ham and chocolate. Mmm... Oh, and perhaps *annoy* my brothers and their girlfriends.

These examples should hopefully help the reader appreciate the difference in usage that Actor type is indicating. Furthermore, notice the difference of register and topic of discourse between example (5) and examples (6) and (7). This is considered in more detail below. One of the most striking features, however, is the presence of humour. It must be remembered that *annoy* is a negative term and its use in a humorous manner is marked.

2.3. Actor Familiarity and Humour

Two other variables that we can consider are the use of humour and familiarity between the Actors. Combining Familiarity and Actor Person and Type results in a degree of multicollinearity that makes Correspondence Analysis unusable, since Familiarity is impossible with Inanimate Actors and effectively universal with Second Person Actors. For this reason, we must consider it separately. Moreover, Humour is difficult to annotate accurately. The presence of Humour was ascribed to the occurrence if and only if there was no doubt that the intent of utterance was humorous. Although this is a subjective criterion, the results prove to be reasonably consistent. Nevertheless, we must consider this variable with due caution. Figure 6 plots a Multiple Correspondence Analysis of these two variables relative to Dialect.

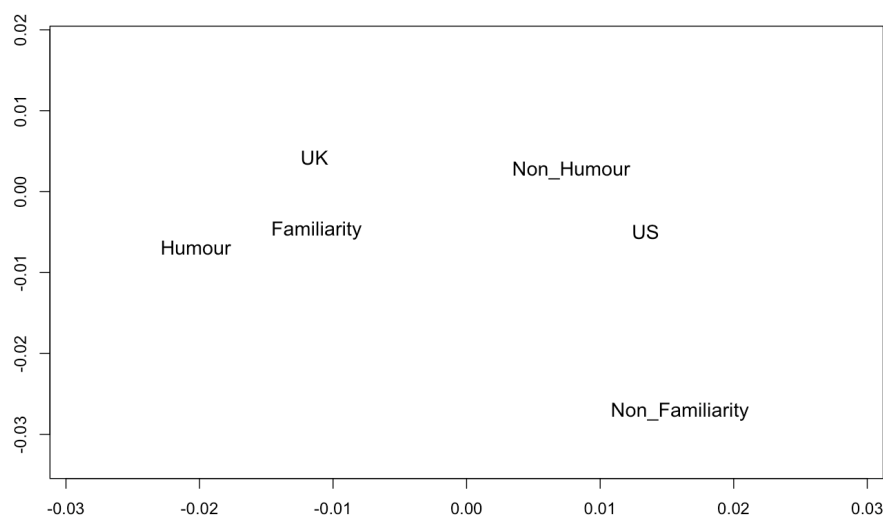


Figure 6. Multiple Correspondence Analysis of *annoy*. Actor Familiarity, Humour, and Dialect

For the Correspondence Analysis presented in Figure 6, the occurrences of inanimate Agents were removed since ‘familiarity’ is impossible in such instances. The analysis clearly shows the correlation between ‘familiarity’ <Familiarity> and British usage <UK> contrasted by ‘non-familiarity’ <Non_Familiarity> between the Actors for the American usage <US>.

Turning to the use of Humour, we see that a ‘lack of humour’ <Non_Humour> is not particularly associated with either of the dialects, but that the ‘presence of humour’ <Humour> is distinctly correlated with the British usage. This correlation between humorous usage and British is statistically significant. In the data set, there are 53 out of 255 UK occurrences of Humour and only 8 out of 256 US occurrences. A binomial test shows that the difference is significant ($p = 2.368e-07$). Examples (8a) - (8b) are representative of the usage of humour in question.

- (8) a. OMG OMG OMG OMG. hahaha waaah so there i was *annoying* the shit out of my mate
- b. when we are meant to be working not *annoying* the librarians with our stupid conversations and non-stop laughing.

Both variables, Humour and Familiarity, seem to distinguish usage between the two dialects. The feature of ‘humour’ is especially important, not only because it is distinctly British in the use, it seems to point to a lighter, less “angry” meaning of *annoy*, at least outside of the canonical use. This less

negative usage may also, hypothetically, explain the association of Actor familiarity with the British usage since it is used in a positive inter-personal manner.

Let us now turn back to the data and see what kind of examples the Correspondence Analysis is identifying. The following four examples are typical of the examples where both Familiarity and Humour are present. There are some 50 examples of this kind in the British data.

- (9) a. [...] a dance mat. I'll have to invite people round to annoy my housemates with it. I suspect minervasolo will want to play LOL.
- b. Sammeth commented that we really shouldn't have wasted so much money when blatantly none of us could sing in key. I quote: "we might as well sit on a fence and yowl, at least we get to annoy neighbors and it's free!"
- c. wel today bin ace coz its bina snowday!!!! jus bin randomly snowballin things an ppl. plus we managed to annoy ma mum by makin a big penis out of snow an putin it on her car...oh the funies.
- d. anyway i was annoying both Fiona and Jade today by prodding them with a pen SHUT UP DIRTY MINDED PEOPLE....

In American, there are only five examples where both Familiarity and Humour are found together. We can list all five examples:

- (9) e. then i walked up and got in bed, where apparently cassidy and i annoyed the crap out of bekah haha cause she had the urge to like talk to jon about us on the phone while we were RIGHT there but ah...h.aha.
- f. Buffy smiled. Hey, you can still annoy Angel from this side, she said brightly. Lindsey winked at her, and she laughed, this time leaning over to meet his lips halfway.
- g. Tom wore black robes with a silver edging of serpents. Harry also, to annoy the Dark Lord, had spiked his hair and turned it green and gold, to match his robes.
- h. So now I gotta go get Rob his birthday present, annoy Pat while he's working at the coffee place, and then hopefully swing by Boomers for some last minute fun.
- i. If I add you and you annoy me, you will have to fear the wrath of my penguins

These examples show that although this grouping of features is typical of the British use of *annoy*, it is not exclusively so. What multifactorial approaches capture are tendencies, not clear-cut divisions. This should be seen as a strength rather than a shortcoming since there is no reason to assume that semasiological variation and conceptual structure should constitute discreet meanings. On the contrary, it seems both linguistically natural and cognitively plausible that sense distinctions necessarily form continua and that discreet differences between meanings are the exception rather than the norm.

2.4. Thematic Topic of Discourse

The variable of Theme, or Topic of Discourse, was mentioned above. An important question concerning this variable is how it is affected by issues of corpus representativity and the interpretation of extra-linguistic variables in semantic analysis. Assuming that the corpus is not skewed and also that the British and Americans normally talk about similar things, we see interesting variation between the dialects for the variable of Theme. The differences in the thematic usage of *annoy* may well be indicative of a difference in meaning or perhaps register. It seems, looking at the variation, that it mirrors some of the differences in the use of Humour and Actor type seen above.

The association plot in Figure 7 captures the thematic differences between the American and British usage. The top half of the plot represents the relative distribution for the British usage and the lower half, the American usage. In this representation, the prominence and equal distribution of the most typical Theme, ‘personal miscellaneous’ <pr_misc>, is clear. Notice, however, the relative importance of the Themes ‘private entertainment’ <pr_ent>, ‘work’ <work>, and ‘school-university’ <uni> in the British examples. A binomial test for comparing two proportions shows that ‘work’ ($p = <2.2e-16$), ‘school-university’ ($p = 2.332e-05$), and ‘private entertainment’ ($p = < 2.2e-16$) are significantly more associated with British usage. The Theme of ‘travel-holidays’ <trav> is also highly associated with the British usage, but the small number of examples of this Theme makes it difficult to draw any conclusions. On the other hand, the Themes of ‘family’ <fam> ($p = 2.2e-16$), ‘personal relationships’ <rel> ($p = 7.299e-11$), and ‘society’ <soc> ($p = 2.2e-16$) are significantly associated with the American data. The Theme of ‘public entertainment’, which covers discussion of media events and celebrities, is distinctly American but the relatively small number of occurrences means this may just be chance.

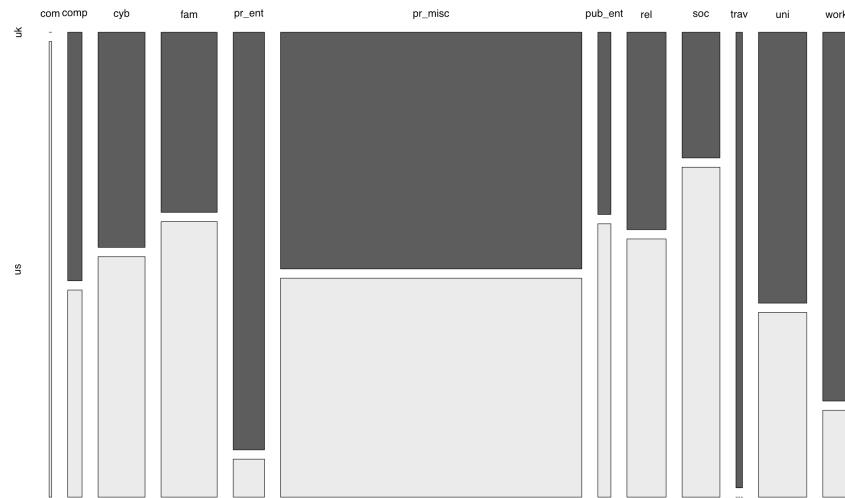


Figure 7. Association plot of *annoy*. Theme and Dialect

We can deduce, therefore, that the American usage tends to be more concerned with emotionally ‘heavy’ subjects such as issues in the family and relationships or society at large, whereas the British usage was more typically talking about going out to the pub and light issues that arise in discussing private entertainment, especially nightlife. Notice also the correlation between British usage and the Themes of ‘work’ <Theme.work>, ‘university–school’ <Theme.uni>, and ‘travel–holidays’ <Theme.trav>. Although not necessarily ‘light’, they are similar to the extent that they are not typically as serious topics as annoyances in the family or in relationships.

These differences in usage are reflected in the plot above. Importantly, they correlate with the findings of the difference in the use of humour. If the use of the term *annoy* is ‘lighter’ in British English, this would explain why it is more likely to be used humorously. Hopefully, it should be clear how these different features of usage come together and their co-occurrence begins to give a semantic picture, but also how that semantic picture differs across the two dialects.

Let us consider some examples captured by this correlation between ‘personal relationships’ <Theme.rel>, ‘family’, <Theme.fam>, ‘social issues’ <Theme.soc> and the American usage of *annoy*. Examples (10a) – (10d) are typical of the ‘family’ and ‘relationship’ Themes. A reasonable amount of context is included to help the reader judge the thematic tendency at stake.

- (10) a. it always makes me so fucking moody. Things that shouldn't get to me are really annoying me. For instance--um I dont know my boyfriend saying he'll see me today then sleeping for 6 fucking hours and having me wait for his call just to go online and have him tell me not tonight.
- b. mom is being really gay about austin which she has been doing a LOT lately and it's really making me upset. basically, to put it in a bubble or whatever, she pretty much makes me feel like i'm not good enough for him and all this junk and that i'm annoying him and his family and all sorts of crap like that. andddd..to top it off, apparently the world thinks i've been a biotch basically to sharon and it totally didn't mean to.
- c. It makes me resent him even though I freely give to him. Im so fucking confused my him. I love him very much, but lately everything he does annoys me. I feel like a mother, nothing gets done in the apartment unless I am upset enough that its worth doing. I am working more than full time and I hate my job,
- d. i hate youuuu....don't get me wrong, i love you...but you annoy the fuck out of me. you hold me back from doing the things i need to do to get on with my life.

Examples (11a) - (11c) are typical of the 'society' Theme. Notice also the seriousness of tone across the examples. Although not all 'society' examples are serious in tone, this tone is typical. Since 'seriousness' is a difficult semantic feature to analyse systematically, thematic features such as these can sometimes serve as indirect indices of such semantic tendencies.

- (11) a. What also annoys me is that whenever you're having a wonderful point in your life, you kind of forget about God. Like, "Well, everything is okay now, so why should I even bother to rely on God?" Then God takes that moment away and you have to rely back on Him
- b. if i didn't know how, or couldn't draw i would die- if i couldn't dance i would die -i like video games... way too much -church utterly annoys me
- c. god some people just think they are sooo cool because they do this or that and they want to tell the whole wide world....it annoys me because people believe stupid people and say shit...

An important aside should be made at this point. In the data, one sees a difference between American and British usage in ‘public entertainment’ <pub_ent> and ‘private entertainment’ <pr_ent> respectively. This, we can hypothesise, is a cultural difference and is not necessarily indicative of difference in lexical meaning. In American culture, there are no public houses and live music is much less common. Indeed, the few private entertainment examples in the American data were restricted to going to the cinema (*movies*) or ‘college parties’. There were, of course, no references to college parties in the British corpus and, although examples concerning going to the cinema do occur, they are offset by a large number of examples discussing a ‘night at the pub’, with or without live music. These, almost stereotypical, differences between the cultures should not lead to an interpretation of difference in the meaning of a single lexeme.

2.5. Cause and Affect

Finally, let us turn to the variables of Cause and Affect. Cause represents the kind of stimulus evoking the state of ANNOY. For example, in *the sight of Jim’s wound annoyed me*, the Agent type would be an inanimate ‘thing’, but the stimulus, or cause, would be the displeasure at seeing something distasteful. Another example would be *doing my nails really annoys me*. In this example, ‘activity’ would be the Agent type but it would be the ‘expenditure of energy’ that is the cause of the ANNOY event. Closely related to this Factor is the Affect experienced by the Patient. Here, the utterances are categorised by the kind of emotion in question. For example, in *his painful remarks annoy me, but I refuse to cry*, we can assume that the annoyance of the Patient is wrapped up with ‘emotional pain’, where in the example *people come up to you on the street and annoy you for money*, we can suppose the Patient feels imposed upon, labelled ‘imposition’. Although such semantic distinctions are difficult to make, with careful reading of large amounts of context, it is possible. This is true at least to the extent that one may distinguish between a few coarse-grain Cause-Affect features.

Figure 8 plots a Multiple Correspondence Analysis of the Cause-Affect variable relative to Dialect, Humour, and Theme. The results show that the Cause-Affect variable coincides with the findings above. The canonical Affect is ‘anger’ <affect.anger>, which lies directly between the two dialects and next to the ‘non-humour’ data point <Non_Humour> and the Theme of ‘private miscellaneous’ <theme.pr_misc> (i).

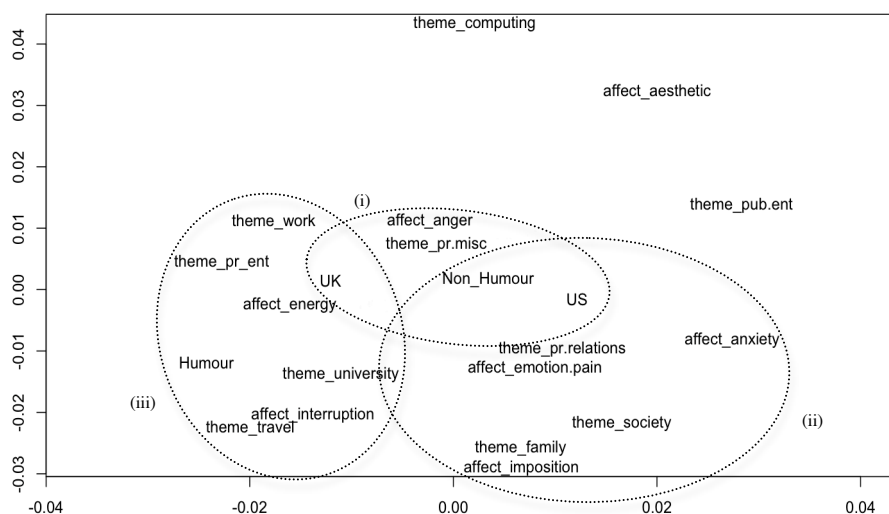


Figure 8. Multiple Correspondence Analysis of *annoy*
Affect, Theme, Humour, and Dialect

On the American side of the plot, we see Cause-Affect type features such as ‘aesthetics’ <affect_aesthetic>, where the cause of the state of ANNOY is something visually or audibly displeasing. More importantly, we see a grouping of the Affects ‘emotional pain’ <affect_emotion.pain> and ‘anxiety’ <affect_anxiety> where the affect experienced by the Patient is emotionally quite serious yet not specifically anger (ii). Lastly, the Affect feature ‘imposition’ <affect_imposition>, where the Patient feels imposed upon by the Agent in the ANNOY event, is also correlated here. In this grouping, these Affect features cluster with the Themes of ‘society’ <theme_society>, ‘personal relations’ <theme_pr.relations>, and ‘family’ <theme_family>. A combination of these Affects and Themes gives a general impression of a serious use of *annoy* where the Patient is hurt, in some way, by the event.

In contrast to the Cause-Affects associated with the American usage, on the British side of the plot we find the feature of ‘expenditure of energy’ <affect_energy> and also, although perhaps less clearly associated, the Cause-Affect of ‘interruption’ <affect_interruption> (iii). These correlations follow the general trend that, outside of the canonical usage, the British usage is less emotionally charged. This trend and the contrast between the dialects are mirrored by the Theme features. Here, for the British usage, we find Themes such as ‘personal entertainment’ <theme_pr.ent>, ‘work’ <theme_work>, ‘school-university’ <theme_university>, and ‘travel’

<theme_travel>. Finally, the distinct correlation of ‘humour’ with the British clustering of usage features, gives a clear picture of a lighter use of *annoy* employed in the discussion of less serious topics.

To finish up, let us briefly consider several examples of these Cause-Affect features of ‘emotional pain’ and ‘anxiety’. These two features were difficult to distinguish in the data but behaved similarly in relation to Dialect, Theme, and Humour. There are some 30 examples of these kinds of occurrences in the American data set compared to only four in the British data. Despite the small numbers, there seems to be a clear trend. Examples (12a) - (12c) are representative of these kinds of occurrences.

- (12) a. If I am to use this journal for what a journal is for, it is like therapy. That means mostly talking about what bothers, plagues, eats away at or annoys me, or just what occupies my mind, really.
- b. It would be great if I could just tell him everything, but alas I can not because of fear. Fear controls me all the time and it annoys me the most of all things. I guess the thing I fear most is fear itself.
- c. Everett is dating a girl from my team. This annoys me on several levels. But the main thing is when I refriended him after the break up because he needed me and had cancer and stuff.

The similarity in tone of examples (14) and examples (12) - (13) should be evident. Such a trend in usage, and a trend that is typical of one dialect and not the other, strongly suggests regionally determined semasiological variation.

3. Summary. Semasiological structure and sociolinguistic effects

In this study, we have examined an approach to polysemy that brings sociolinguistic parameters of language structure into the main of semantic analysis. We have shown that multifactorial techniques, identifying patterns of usage relative to a range of factors, can indentify ‘senses’ while nuancing these senses with sociolinguistic information such as variation in regional use.

In order to appreciate the importance of this, we can consider the findings contrasted with more traditional semantic descriptions. The American Webster's definition distinguished two senses for the transitive form of *annoy*: (i) to disturb or irritate especially by repeated acts and (ii) to harass

especially by quick brief attacks. The Oxford English Dictionary also gives two current 'meanings': (i) cause slight anger or mental distress and (ii) molest, injure, harass. These definitions do not correlate with our findings. Indeed, the iterative Aktionsart suggested in the American definitions was completely absent in the data set. However, the definition 'cause slight anger or mental distress' might be a good summary of the canonical usage we identified. Yet then, the definition 'molest, injure, harass' does not correspond at all to the examples in the corpus. The dictionaries also list intransitive meanings, yet the intransitive form of the verb was entirely absent in a data set of some 750 examples. This, of course, could just be the result of a relatively small sample, but may also indicate that the intransitive form is less typical of the verb or that it is an older form.

Summarised in a way that resembles dictionary entries, this study revealed three basic *senses* of the verb *annoy*:

annoy -

A. [*anger annoy*]

Canonical Usage, typical of both dialects

- A familiar third party angers the patient. It is non-humorous and is typically caused by some personal issue. In American only, this is also associated with an Agent in the second person. This second person usage is often confrontational in nature suggesting that it falls between the canonical [*anger*] usage and the [*hurt annoy*] usage, typical of American.

B. [*hurt annoy*]

Typical of American

- A third party hurts or upsets the patient. This usage is associated with serious Affect, such as emotional stress-anxiety. Thematically, it is used when discussing relatively serious topics such as personal relationship troubles or family concerns. It is also employed in discussion of social issues at large.

C. [*tire-interrupt annoy*]

Typical of British

- The Agent is typically a person or an event in the third person. In the case of human Agents, there is often familiarity between the Agent and the Patient. It is associated with trivial causes such as interruptions and the expenditure of energy. It is often used humorously.

We should, of course, investigate further. There are many other possibilities to consider, for example, the interaction of Aspect with grammatical negation and Affect or the interaction of these with Actor Type and Dialect. However, it has been demonstrated how quantitative data can be used to study semantic variation while accounting for extra-linguistic factors. Specifically, we saw how the variable dialect plays a crucial role in lexical polysemy. Such subtle, yet important, differences in regional variation must be incorporated into the main of Cognitive Semantics.

Notes

- 1 This study was completed with a grant from the University of Leuven. I would like to thank my colleagues there, especially but not exclusively, Koen Plevoets for showing me Correspondence Analysis and Sofie van Gijssel for her time and patience in teaching me R. I am equally indebted to Dirk Geeraerts and Dirk Speelman for all their help and advice.
- 2 Geeraerts (2006) and Glynn (2010c) give an overview of this trend in Cognitive Linguistics.

References

- Croft, W. (2008). Toward a social cognitive linguistics. In V. Evans & S. Pourcel (Eds.), *New Directions in Cognitive Linguistics* (pp. 395–420). Amsterdam: John Benjamins.
- Cuyckens, H. (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* (pp. 27-72). Berlin: Mouton de Gruyter.
- Cuyckens, H. (1995). Family Resemblance in the Dutch Spatial Prepositions *Door* and *Langs*. *Cognitive Linguistics*, 6, 183-207.
- Deane, P. (1998). Polysemy and cognition. *Lingua*, 75, 325-361.
- Deane, P. (2005). Multimodal spatial representation: On the semantic unity of *over*. In B. Hampe (Ed.), *From Perception to Meaning. Image schemas in Cognitive Linguistics* (pp. 235-82). Berlin: Mouton de Gruyter.
- Dirven, R., Goossens, L., Putseys, Y. & Vorlat, E. (1982). *The scene of linguistic action and its perspectivization by SPEAK, TALK, SAY, and TELL*. Amsterdam: John Benjamins.
- Divjak, D. (2006). Ways of Intending: Delineating and structuring near-synonyms. In St. Th. Gries & A. Stefanowitsch (Eds.), *Corpora in Cognitive Linguistics: Corpus-based approaches to syntax and lexis* (pp. 19-56). Berlin: Mouton de Gruyter.

- Divjak, D. (2010) *Structuring the Lexicon: A clustered model for near-synonymy*. Berlin: Mouton de Gruyter.
- Evans, V. (2004). Spatial experience, lexical structure, and motivation: The case of *in*. In G. Radden & K.-U. Panther (ed.), *Linguistics Studies in Motivation* (pp. 157-192). Berlin: Mouton de Gruyter
- Evans, V. (2006). Lexical concepts, cognitive models and meaning-construction. *Cognitive Linguistics*, 17, 491-534.
- Geeraerts, D. (Ed.). (1989). *Prospects and Problems of Prototype Theory* (Special issue of *Linguistics*, 27). Berlin: Mouton de Gruyter.
- Geeraerts, D. (2005). Lectal data and empirical variation in Cognitive Linguistics. In F. J. Ruiz de Mendoza Ibañez and S. Peña Cervel (Eds.), *Cognitive Linguistics. Internal dynamics and interdisciplinary interactions* (pp. 163-189). Berlin: Mouton de Gruyter.
- Geeraerts, D. (2006). Methodology in Cognitive Linguistics. In G. Kristiansen, M. Achard, R. Dirven & F. J. Ruiz de Mendoza Ibañez (Eds.), *Cognitive Linguistics: Current applications and future perspectives* (pp. 21-50). Berlin: Mouton de Gruyter.
- Geeraerts, D., Kristiansen, G. & Peirsman, Yves (Eds.). (2010). *Advances in Cognitive Sociolinguistics*. Berlin / New York: Mouton de Gruyter.
- Gibbs, R. (2007). Why cognitive linguists should care more about empirical methods. In M. Gonzalez-Marquez, I. Mittleberg, S. Coulson, & M. Spivey (Eds.), *Methods in Cognitive Linguistics* (pp. 2-18). Amsterdam: John Benjamins.
- Gilquin, G. (2010). *Corpus, Cognition and Causative Constructions*. Amsterdam: John Benjamins.
- Gilquin, G. (Ed.). (2009). *Corpora and Experimental Methods* (Special ed. *Corpus Linguistics and Linguistic Theory*, 5). Berlin: Mouton de Gruyter.
- Glynn, D. (2003). Polysemy and over-polysemy. A three tiered approach to sense variation. Paper presented at the *8th International Cognitive Linguistics Conference*. Lorogño, Spain.
- Glynn, D. (2009). Polysemy, syntax, and variation. A usage-based method for Cognitive Semantics. In V. Evans & S. Pourcel (Eds.), *New Directions in Cognitive Linguistics* (pp. 77-104). Amsterdam: John Benjamins.
- Glynn, D. (2010a). Synonymy, lexical fields, and grammatical constructions. A study in usage-based Cognitive Semantics. In H.-J. Schmid & S. Handl (Eds.), *Cognitive Foundations of Linguistic Usage-Patterns* (pp. 89-118). Berlin: Mouton de Gruyter.
- Glynn, D. (2010b). Testing the Hypothesis. Objectivity and Verification in Usage-Based Cognitive Semantics. In D. Glynn & K. Fischer (Eds.), *Quantitative Cognitive Semantics. Corpus-driven approaches* (pp. 239-270). Berlin: Mouton de Gruyter.
- Glynn, D. (2010c). Corpus-Driven Cognitive Semantics. An introduction to the field. In D. Glynn & K. Fischer (Eds.), *Quantitative Cognitive Semantics. Corpus-driven approaches* (pp. 1-42). Berlin: Mouton de Gruyter.

- Glynn, D. (in press). Correspondence Analysis. Hunting for patterns of correlation. D. Glynn & J. Robinson (Eds.), *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*. Amsterdam: John Benjamins.
- Glynn, D. & Fischer, K. (Eds.). (2010). *Corpus-driven Cognitive Semantics. Quantitative approaches*. Berlin: Mouton de Gruyter.
- Glynn, D. & Robinson, J. (Eds.). (in press). *Polysemy and Synonymy. Corpus methods and applications in Cognitive Linguistics*. Amsterdam: John Benjamins.
- Gries, Stefan Th. (2003). *Multifactorial Analysis in Corpus Linguistics: A study of particle placement*. London: Continuum Press.
- Gries, St. Th. (2006). Corpus-based methods and cognitive semantics: the many meanings of *to run*. In St. Th. Gries and A. Stefanowitsch (Eds.), *Corpora in Cognitive Linguistics: Corpus-based approaches to syntax and lexis* (pp. 57-99). Berlin: Mouton de Gruyter.
- Gries, St. Th. & David, C. (2007). This is *kind of/sort of interesting*: Variation in hedging in English. In Pahta, Päivi, Irma Taavitsainen, Terttu Nevalainen, & Jukka Tyrkkoö (Eds.). *Towards multimedia in corpus linguistics. Studies in variation, contacts and change in English 2*, University of Helsinki.
- Gries, St. Th. & Stefanowitsch, A. (Eds.). (2006). *Corpora in Cognitive Linguistics: Corpus-based approaches to syntax and lexis*. Berlin: Mouton de Gruyter.
- Heylen, K., Tummers, J. & Geeraerts, D. (2008). Methodological issues in corpus-based Cognitive Linguistics. In G. Kristiansen & R. Dirven (Eds.), *Cognitive Sociolinguistics: Language variation, cultural models, social systems* (pp. 91-128). Berlin: Mouton de Gruyter.
- Hilpert, M. (2008). *Germanic Future Constructions. A usage-based approach to language change*. Amsterdam: John Benjamins.
- Kristiansen, G. & Dirven, R. (Eds.). (2008). *Cognitive Sociolinguistics: Language variation, cultural models, social systems*. Berlin: Mouton de Gruyter.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things. What categories reveal about the mind*. London: University of Chicago Press.
- Langacker, R. (1987). *Foundations of Cognitive Grammar. Vol. 1. Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, R. (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics* (pp. 127-161) Amsterdam: John Benjamins.
- Sandra, D. & Rice, S. (1995). Network analyses of prepositional meaning: Mirroring whose mind - the linguist's or the language user's? *Cognitive Linguistics*, 6, 89-130.
- Schmid, H.-J. (2000). *English Abstract Nouns as Conceptual Shells. From corpus to cognition*. Berlin: Mouton de Gruyter.
- Schmid, H.-J. (2010). Does frequency in text instantiate entrenchment in the cognitive system? *Quantitative Cognitive Semantics: Corpus-driven approaches*, D. Glynn & K. Fischer (Eds.), 101-135. Berlin: Mouton de Gruyter.

- Speelman, D. & Geeraerts, D. (2010). Causes for causatives: the case of Dutch 'doen' and 'laten'. In T. Sanders & E. Sweetser (Eds.), *Causal Categories in Discourse and Cognition* (pp. 173-204). Berlin: Mouton de Gruyter.
- Stefanowitsch, A. & Gries, St. Th. (2006). *Corpus-Based Approaches to Metaphor and Metonymy*. Berlin: Mouton de Gruyter.
- Tummers, J., Heylen, K. & Geeraerts, D. (2005). Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory*, 1, 225-261.
- Rudzka-Ostyn, B. (1989). Prototypes, schemas, and cross-category correspondences: The case of *ask*. In D. Geeraerts (Ed.), *Prospects and Problems of Prototype Theory* (Special issue of *Linguistics*, 27) (pp. 613-61). Berlin: Mouton de Gruyter.
- Rudzka-Ostyn, B. (1995). Metaphor, schema, invariance: The case of verbs of answering. In L. Goossens, P. Pauwels, B. Rudzka-Ostyn, A.-M. Simon-Vandenberghe & J. Vanparys (Eds.), *By Word of Mouth. Metaphor, metonymy, and linguistic action from a cognitive perspective* (pp. 205-244). Amsterdam: John Benjamins.
- Vandeloise, C. (1985). *L'espace en français*. Paris: Seuil.
- Verschueren, J. (1985). *What People Say They do with Words: Prolegomena to an empirical-conceptual approach to linguistic action*. Norwood: Ablex.
- Wulff, S. (2009). *Rethinking Idiomaticity. A usage-based approach*. London: Continuum.
- Wulff, S., Stefanowitsch, A. & Gries, St. Th. (2007). Brutal Brits and persuasive Americans: Variety-specific meaning construction in the into-causative. In G. Radden, K.-M. Köpcke, T. Berg & P. Siemund (Eds.), *Aspects of Meaning Construction* (pp. 265-281). Amsterdam: John Benjamins.
- Zelinsky-Wibbelt, C. (2000). *Discourse and the Continuity of Reference. Representing mental categorization*. Berlin: Mouton de Gruyter.
- Zeschel, A. (Ed.) (2008). *Usage-based Approaches to Language Processing and Representation* (Special ed. *Cognitive Linguistics*, 19). Berlin: Mouton de Gruyter.