Semantic Density Analysis: Comparing word meaning across time and phonetic space

Eyal Sagi
Northwestern University
Evanston, Illinois, USA
eyal@u.northwestern.edu

Stefan Kaufmann
Northwestern University
Evanston, Illinois, USA
kaufmann@northwestern.edu

Brady Clark
Northwestern University
Evanston, Illinois, USA
bzack@northwestern.edu

Abstract

This paper presents a new statistical method for detecting and tracking changes in word meaning, based on Latent Semantic Analysis. By comparing the density of semantic vector clusters this method allows researchers to make statistical inferences on questions such as whether the meaning of a word changed across time or if a phonetic cluster is associated with a specific meaning. Possible applications of this method are then illustrated in tracing the semantic change of 'dog', 'do', and 'deer' in early English and examining and comparing phonaesthemes.

1 Introduction

The increase in available computing power over the last few decades has led to an explosion in the application of statistical methods to the analysis of texts. Researchers have applied these methods to a wide range of tasks, from word-sense disambiguation (Levin et al., 2006) to the summarization of texts (Marcu, 2003) and the automatic scoring of student essays (Riedel et al., 2006). However, some fields of linguistics that have traditionally employed corpora as their source material, such as historical semantics, have yet to benefit from the application of these statistical methods.

In this paper we demonstrate how an existing statistical tool (Latent Semantic Analysis) can be adapted and used to automate and enhance some aspects of research in historical semantics and other fields whose focus is on the comparative analysis of word meanings within a corpus. Our method allows us to assess the semantic variation within the set of individual occurrences of a given word type. This variation is inversely related to a property of types that we call *density* – intuitively, a tendency to occur in highly similar contexts. In terms of our LSA-based spatial semantic model, we calculate vectors representing the context of each occurrence of a given term, and estimate the term's cohesiveness as the density with which these token context vectors are "packed" in space.

2 The method

Latent Semantic Analysis (LSA) is a collective term for a family of related methods, all of which involve building numerical representations of words based on occurrence patterns in a training corpus. The basic underlying assumption is that co-occurrence within the same contexts can be used as a stand-in measure of semantic relatedness (see Firth, 1957; Halliday and Hasan, 1976; Hoey, 1991, for early articulations of this idea). The success of the method in technical applications such as information retrieval and its popularity as a research tool in psychology, education, linguistics and other disciplines suggest that this hypothesis holds up well for the purposes of those applications.

The relevant notion of "context" varies. The first and still widely used implementation of the idea, developed in Information Retrieval and originally known as Latent Semantic Indexing (Deerwester et al., 1990), assembles a term-document matrix in which each vocabulary item (term) is associated with an n-dimensional vector recording its distribution over the n documents in the corpus. In contrast, the version we applied in this work measures co-occurrence in a way that is more independent of the characteristics of the documents in the training corpus, building in-

Proceedings of the EACL 2009 Workshop on GEMS: GEometical Models of Natural Language Semantics, pages 104–111,
Athens, Greece, 31 March 2009.

stead a term-term matrix associating vocabulary items with vectors representing their frequency of co-occurrence with each of a list of "content-bearing" words. This approach originated with the "WordSpace" paradigm developed by Schütze (1996). The software we used is a version of the "Infomap" package developed at Stanford University and freely available (see also Takayama et al., 1999). We describe it and the steps we took in our experiments in some detail below.

2.1 Word vectors

The information encoded in the co-occurrence matrix, and thus ultimately the similarity measure depends greatly on the genre and subject matter of the training corpus (Takayama et al., 1999; Kaufmann, 2000). In our case, we used the entire available corpus as our training corpus. The word types in the training corpus are ranked by frequency of occurrence, and the Infomap system automatically selects (i) a vocabulary W for which vector representations are to be collected, and (ii) a set C of 1,000 "content-bearing" words whose occurrence or non-occurrence is taken to be indicative of the subject matter of a given passage of text. Usually, these choices are guided by a stoplist of (mostly closed-class) lexical items that are to be excluded, but because we were interested in tracing changes in the meaning of lexical items we reduced this stoplist to a bare minimum. To compensate, we increased the number of "content-bearing" words to 2,000. The vocabulary W consisted of the 40,000 most frequent non-stoplist words. The set C of content-bearing words contained the 50th through 2,049th most frequent non-stoplist words. This method may seem rather blunt, but it has the advantage of not requiring any human intervention or antecedently given information about the domain.

The cells in the resulting matrix of 40,000 rows and 2,000 columns were filled with co-occurrence counts recording, for each pair $\langle w, c\rangle \in W \times C$, the number of times a token of c occurred in the context of a token of w in the corpus.[1] The "context" of a token w_i in our implementation is the set of tokens in a fixed-width window from the 15th item preceding w_i to the 15th item following it (less if a document boundary intervenes). The matrix was transformed by Singular Value Decomposition (SVD), whose implementation in the Infomap system relies on the SVDPACKC package (Berry, 1992; Berry et al., 1993). The output was a reduced 40,000 × 100 matrix. Thus each item $w \in W$ is associated with a 100-dimensional vector $\vec{w}$.

2.2 Context vectors

Once the vector space is obtained from the training corpus, vectors can be calculated for any multi-word unit of text (e.g. paragraphs, queries, or documents), regardless of whether it occurs in the original training corpus or not, as the normalized sum of the vectors associated with the words it contains. In this way, for each occurrence of a target word type under investigation, we calculated a *context vector* from the 15 words preceding and the 15 words following that occurrence.

Context vectors were first used in Word Sense Discrimination by Schütze (1998). Similarly to that application, we assume that these "second-order" vectors encode the aggregate meaning, or topic, of the segment they represent, and thus, following the reasoning behind LSA, are indicative of the meaning with which it is being used on that particular occurrence. Consequently, for each target word of interest, the context vectors associated with its occurrences constitute the data points. The analysis is then a matter of grouping these data points according to some criterion (e.g., the period in which the text was written) and conducting an appropriate statistical test. In some cases it might also be possible to use regression or apply a clustering analysis.

2.3 Semantic Density Analysis

Conducting statistical tests comparing groups of vectors is not trivial. Fortunately, some questions can be answered based on the similarity of vectors within each group rather than the vectors themselves. The similarity between two vectors $\vec{w}$, $\vec{v}$ is measured as the cosine between them:[2]

[1] Two details are glossed over here: First, the Infomap system weighs this raw count with a $tf.idf$ measure of the column label c, calculated as follows: $tf.idf(c) = tf(c) \times \left(log(D+1) - log(df(c))\right)$ where tf and df are the number of occurrences of c and the number of documents in which c occurs, respectively, and D is the total number of documents. Second, the number in each cell is replaced with its square root, in order to approximate a normal distribution of counts and attenuate the potentially distorting influence of high base frequencies (cf. Takayama, et al. 1998; Widdows, 2004).

[2] While the cosine measure is the accepted measure of similarity, the cosine function is non-linear and therefore problematic for many statistical methods. Several transformations can be used to correct this (e.g., Fisher's z). In this paper we will use the angle, in degrees, between the two vectors (i.e., cos^{-1}) because it is easily interpretable.

$$cos(\vec{w}, \vec{v}) = \frac{\vec{w} \cdot \vec{v}}{|\vec{w}||\vec{v}|}$$

The average similarity of a group of vectors is indicative of its density – a dense group of highly similar vectors will have a high average cosine (and a correspondingly low average angle) whereas a sparse group of dissimilar vectors will have an average cosine that approaches zero (and a correspondingly high average angle).[3] Thus since a word that has a single, highly restricted meaning (e.g. 'palindrome') is likely to occur in a very restricted set of contexts, its context vectors are also likely to have a low average angle between them, compared to a word that is highly polysemous or appears in a large variety of contexts (e.g. 'bank', 'do'). From this observation, it follows that it should be possible to compare the cohesiveness of groups of vectors in terms of the average pairwise similarity of the vectors of which they are comprised. Because the number of such pairings tends to be prohibitively large (e.g., nearly 1,000,000 for a group of 1,000 vectors), it is useful to use only a sub-sample in any single analysis. A Monte-Carlo analysis in which n pair-wise similarity values are chosen at random from each group of vectors is therefore appropriate.[4]

However, there is one final complication to consider in the analysis. The passage of time influences not only the meaning of words, but also styles and variety of writing. For example, texts in the 11th century were much less varied, on average, than those written in the 15th century.[5] This will influence the calculation of context vectors as those depend, in part, on the text they are taken from. Because the document as a whole is represented by a vector that is the average of all of its words, it is possible to predict that, if no other factors exist, two contexts are likely to be related to one another to the same degree that their documents are. Controlling for this effect can therefore be achieved by subtracting from the angle between two context vectors the angle between the documents in which they appear.

3 Applications to Research

3.1 A Diachronic Investigation: Semantic Change

One of the central questions of historical semantics is the following (Traugott, 1999):[6]

> Given the form-meaning pair L (lexeme) what changes did meaning M undergo?

For example, the form *as long as* underwent the change `equal in length' > `equal in time' > `provided that'. Evidence for semantic change comes from written records, cognates, and structural analysis (Bloomfield, 1933). Traditional categories of semantic change include (Traugott, 2005: 2-4; Campbell, 2004:254-262; Forston, 2003: 648-650):

- Broadening (generalization, extension, borrowing): A restricted meaning becomes less restricted (e.g. Late Old English *docga* `a (specific) powerful breed of dog' > *dog* `any member of the species *Canis familiaris*'
- Narrowing (specialization, restriction): A relatively general meaning becomes more specific (e.g. Old English *deor* `animal' > *deer*)
- Pejoration (degeneration): A meaning becomes more negative (e.g. Old English *sælig* `blessed, blissful' > *sely* `happy, innocent, pitiable' > *silly* `foolish, stupid')

Semantic change results from the use of language in context, whether linguistic or extralinguistic. Later meanings of forms are connected to earlier ones, where all semantic change arises by polysemy, i.e. new meanings coexist with earlier ones, typically in restricted contexts. Sometimes new meanings split off from earlier ones and are no longer considered variants by language users (e.g. *mistress* `woman in a position of authority, head of household' > `woman in a continuing extra-marital relationship with a man').

Semantic change is often considered unsystematic (Hock and Joseph, 1996: 252). However, recent work (Traugott and Dasher, 2002) suggests that there is, in fact, significant cross-linguistic regularity in semantic change. For ex-

[3] Since the cosine ranges from -1 to +1, it is possible in principle to obtain negative average cosines. In practice, however, the overwhelming majority of vocabulary items have a non-negative cosine with any given target word, hence the average cosine usually does not fall below zero.

[4] It is important to note that the number of *independent samples* in the analysis is determined not by the number of similarity values compared but by the number of individual vectors used in the analysis.

[5] Tracking changes in the distribution of the document vectors in a corpus over time might itself be of interest to some researchers but is beyond the scope of the current paper.

[6] This is the *semasiological* perspective on semantic change. Other perspectives include the *onomasiological* perspective ("Given the concept C, what lexemes can it be expressed by?"). See Traugott 1999 for discussion.

Table 1 - Mean angle between context vectors for target words in different periods in the Helsinki corpus (standard deviations are given in parenthesis)

	n	*Unknown composition date (<1250)*	*Early Middle English (1150-1350)*	*Late Middle English (1350-1500)*	*Early Modern English (1500-1710)*
dog	112			15.47 (14.19)	24.73(10.43)
do	4298		10.31(13.57)	13.02 (9.50)	24.54 (11.2)
deer	61	38.72 (17.59)	20.6 (18.18)		20.5 (9.82)
science	79			13.56 (13.33)	28.31 (12.24)

ample, in the Invited Inferencing Model of Semantic Change proposed by Traugott and Dasher (2002) the main mechanism of semantic change is argued to be the semanticization of *conversational implicatures*, where conversational implicatures are a component of speaker meaning that arises from the interaction between what the speaker says and rational principles of communication (Grice, 1989 [1975]). Conversational implicatures are suggested by an utterance but not entailed. For example, the utterance *Some students came to the party* strongly suggests that some but not all students came to the party, even though the utterance would be true strictly speaking if all students came to the party. According to the Invited Inferencing Model, conversational implicatures become part of the semantic polysemies of particular forms over time.

Such changes in meaning should be evident when examining the contexts in which the lexeme of interest appears. In other words, changes in the meaning of a type should translate to differences in the contexts in which its tokens are used. For instance, semantic broadening results in a meaning that is less restricted and as a result can be used in a larger variety of contexts. In a semantic space that encompasses the period of such a change, this increase in variety can be measured as a decrease in vector density across the time span of the corpus. This decrease translates into an increase in the average angle between the context vectors for the word. For instance, because the Old English word ‘docga’ applied to a specific breed of dog, we predicted that earlier occurrences of the lexemes ‘docga’ and ‘dog’, in a corpus of documents of the appropriate time period, will show less variety than later occurrences.

An even more extreme case of semantic broadening is predicted to occur as part of the process of grammaticalization (Traugot and Dasher, 2002) in which a content word becomes a function word. Because, as a general rule, a function word can be used in a much larger variety of contexts than a content word, a word that underwent grammaticalization should appear in a substantially larger variety of contexts than it did prior to becoming a function word. One well studied case of grammaticalization is that of periphrastic ‘do’. While in Old English ‘do’ was used as a verb with a causative and habitual sense (e.g. ‘do you harm’), later in English it took on a functional role that is nearly devoid of meaning (e.g. ‘do you know him?’). Because this change occurred in Middle English, we predicted that earlier occurrences of ‘do’ will show less variety than later ones.

In contrast with broadening, semantic narrowing results in a meaning that is more restricted, and is therefore applicable in fewer contexts than before. This decrease in variety results in an increase in vector density and can be directly measured as a decrease in the average angle between the context vectors for the word. As an example, the Old English word ‘deor’ denoted a larger group of living creatures than does the Modern English word ‘deer’. We therefore predicted that earlier occurrences of the lexemes ‘deor’ and ‘deer’, in a corpus of the appropriate time period, will show more variety than later occurrences.

We tested our predictions using a corpus derived from the Helsinki corpus (Rissanen, 1994). The Helsinki corpus is comprised of texts spanning the periods of Old English (prior to 1150A.D.), Middle English (1150-1500A.D.), and Early Modern English (1500-1710A.D.). Because spelling in Old English was highly variable, we decided to exclude that part of the corpus and focused our analysis on the Middle English and Early Modern English periods. The resulting corpus included 504 distinct documents totaling approximately 1.1 million words.

To test our predictions regarding semantic change in the words ‘dog’, ‘do’, and ‘deer’, we collected all of the contexts in which they appear in our subset of the Helsinki corpus. This resulted in 112 contexts for ‘dog’, 4298 contexts for ‘do’, and 61 contexts for ‘deer’. Because there were relatively few occurrences of ‘dog’

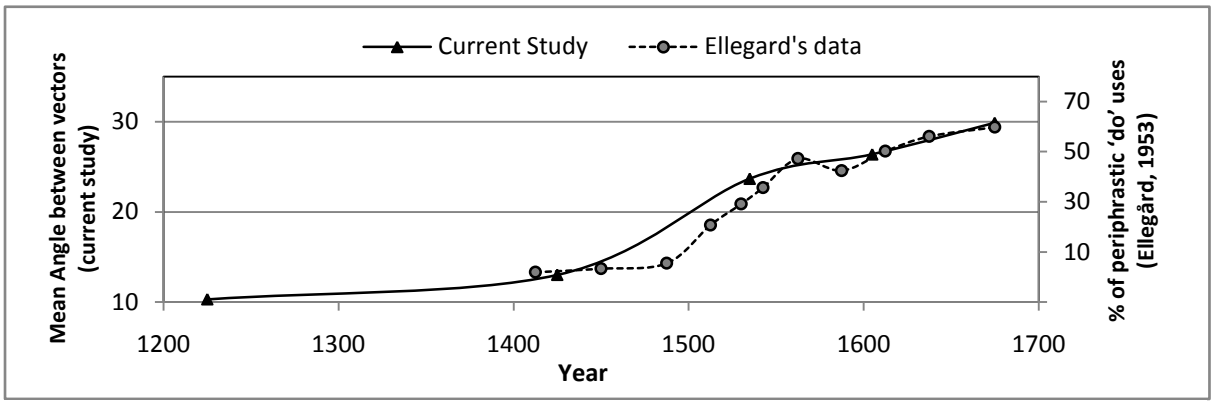

Figure 1 – A comparison of the rise of periphrastic 'do' as measured by semantic density in our study and the proportion of periphrastic 'do' uses by Ellegård (1953).

and 'deer' in the corpus it was practical to compute the angles between all possible pairs of context vectors. As a result, we elected to forgo the Monte-Carlo analysis for those two words in favor of a full analysis. The results of our analysis for all three words are given in Table 1. These results were congruent with our prediction: The density of the contexts decreases over time for both 'dog' ($t(110) = 2.17$, $p < .05$) and 'do' ($F(2,2997)=409.41$, $p < .01$) while in the case of 'deer' there is an increase in the density of the contexts over time ($t(36) = 3.05$, $p < .01$).

Furthermore, our analysis corresponds with the data collected by Ellegård (1953). Ellegård traced the grammaticalization of 'do' by manually examining changes in the proportions of its various uses between 1400 and 1700. His data identifies an overall shift in the pattern of use that occurred mainly between 1475 and 1575. Our analysis identifies a similar shift in patterns between the time periods spanning 1350-1500 and 1500-1570. Figure 1 depicts an overlay of both datasets. The relative scale of the two sets was set so that the proportions of 'do' uses at 1400 and 1700 (the beginning and end of Ellegård's data, respectively) match the semantic density measured by our method at those times.

Finally, our method can be used not only to test predictions based on established cases of semantic change, but also to identify new ones. For instance, in examining the contexts of the word 'science' we can identify that it underwent semantic broadening shortly after it first appeared in the 14th century ($t(77) = 4.51$, $p < .01$). A subsequent examination of the contexts in which the word appears indicated that this is probably the result of a shift from a meaning related to generalized knowledge (e.g., '…and learn science of school', John of Trevisa's Polychronicon, 1387) to one that can also be used to refer to more specific disciplines (e.g., '…of the seven liberal sciences', Simon Forman's Diary, 1602).

Our long term goal with respect to this type of analysis is to use this method in a computer-based tool that can scan a diachronic corpus and automatically identify probable cases of semantic change within it. Researchers can then use these results to focus on identifying the specifics of such changes, as well as examine the overall patterns of change that exist in the corpus. It is our belief that such a use will enable a more rigorous testing and refinement of existing theories of semantic change.

3.2 A Synchronic Investigation: Phonaesthemes

In addition to examining changes in meaning across time, it is also possible to employ our method to examine how the semantic space relates to other possible partitioning of the lexemes represented by it. For instance, while the relationship between the phonetic representation and semantic content is largely considered to be arbitrary, there are some notable exceptions. One interesting case is that of *phonaesthemes* (Firth, 1930), sub-morphemic units that have a predictable effect on the meaning of the word as a whole. In English, one of the more frequently mentioned phonaesthemes is a word-initial *gl-* which is common in words related to the visual modality (e.g., 'glance', 'gleam'). While there have been some scholastic explorations of these non-morphological relationships between sound and meaning, they have not been thoroughly explored by behavioral and computational research (with some notable exceptions; e.g. Hutchins, 1998; Bergen, 2004). Recently, Otis and Sagi (2008) used the semantic density of the cluster of words sharing a phonaestheme as a measure of

the strength of the relationship between the phonetic cluster and its proposed meaning.

Otis and Sagi used a corpus derived from Project Gutenberg (http://www.gutenberg.org/) as the basis for their analysis. Specifically, they used the bulk of the English language literary works available through the project's website. This resulted in a corpus of 4034 separate documents consisting of over 290 million words.

The bulk of the candidate phonaesthemes they tested were taken from the list used by Hutchins (1998), with the addition of two candidate phonaesthemes (*kn-* and *-ign*). Two letter combinations that were considered unlikely to be phonaesthemes (*br-* and *z-*) were also included in order to test the method's capacity for discriminating between phonaesthemes and non-phonaesthemes. Overall Otis and Sagi (2008) examined 47 possible phonaesthemes.

In cases where a phonetic cluster represents a phonaestheme, it intuitively follows that pairs of words sharing that phonetic cluster are more likely to share some aspect of their meaning than pairs of words chosen at random. Otis and Sagi tested whether this was true for any specific candidate phonaestheme using a Monte-Carlo analysis. First they identified all of the words in the corpus sharing a conjectured phonaestheme[7] and chose the most frequent representative word form for each stem, resulting in a cluster of word types representing each candidate phonaestheme.[8] Next they tested the statistical significance of this relationship by running 100 *t*-test comparisons. Each of these tests compared the relationship of 50 pairs of words chosen at random from the conjectured cluster with 50 pairs of words chosen at random from a similarly sized cluster, randomly generated from the entire corpus. The number of times these *t*-tests resulted in a statistically significant difference ($\alpha = .05$) was recorded. This analysis was repeated 3 times for each conjectured phonaestheme and the median value was used as the final result.

To determine whether a conjectured phonaestheme was statistically supported by their analysis Otis and Sagi compared the overall frequency of statistically significant t-tests with the binomial distribution for their α (.05). After applying a Bonferroni correction for performing 50 comparisons, the threshold for statistical significance of the binomial test was for 14 t-tests out of 100 to turn out as significant, with a frequency of 13 being marginally significant. Therefore, if the significance frequency (*#Sig* below) of a candidate phonaestheme was 15 or higher, that phonaestheme was judged as being supported by statistical evidence. Significance frequencies of 13 and 14 were considered as indicative of a phonaestheme for which there was only marginal statistical support.

Among Hutchins' original list of 44 possible phonaesthemes, 26 were found to be statistically reliable and 2 were marginally reliable. Overall the results were in line with the empirical data collected by Hutchins. By way of comparing the two datasets, *#Sig* and Hutchins' average rating measure were well correlated ($r = .53$). Neither of the unlikely phonaestheme candidates we examined were statistically supported phonaesthemes ($\#Sig_{\text{br-}} = 6$; $\#Sig_{\text{z-}} = 5$), whereas both of our newly hypothesized phonaesthemes were statistically supported ($\#Sig_{\text{kn-}} = 28$; $\#Sig_{\text{-ign}} = 23$). In addition to being able to use this measure as a decision criterion as to whether a specific phonetic cluster might be phonaesthemic, it can also be used to compare the relative strength of two such clusters. For instance, in the Gutenberg corpus the phonaesthemic ending *–owl* (e.g., 'growl', 'howl'; *#Sig*=97) was comprised of a cluster of words that were more similar to one another than *–oop* (e.g., 'hoop', 'loop'; *#Sig*=32).

Such results can then be used to test the cognitive effects of phonaesthemes. For instance, following the comparison above, we might hypothesize that the word 'growl' might be a better semantic prime for 'howl' than the word 'hoop' is for the word 'loop'. In contrast, because a word-initial *br-* is not phonaesthemic, the word 'breeze' is unlikely to be a semantic prime for the word 'brick'. In addition, it might be interesting to combine the diachronic analysis from the previous section with the synchronic analysis in this section to investigate questions such as when and how phonaesthemes come to be part of a language and what factors might affect the strength of a phonaestheme.

4 Discussion

While the method presented in this paper is aimed towards quantifying semantic relation-

[7] It is important to note that due to the nature of a written corpus, the match was orthographical rather than phonetic. However, in most cases the two are highly congruent.

[8] Because, in this case, Otis and Sagi were not interested in temporal changes in meaning, they used the overall word vectors rather than look at each context individually. As a result, each of the vectors used in the analysis is based on occurrences in many different documents and there was no need to control for the variability of the documents.

ships that were previously difficult to quantify, it also raises an interesting theoretical issue, namely the relationship between the statistically computed semantic space and the actual semantic content of words. On the one hand, simulations based on Latent Semantic Analysis have been shown to correlate with cognitive factors such as the acquisition of vocabulary and the categorization of texts (cf. Landauer & Dumais, 1997). On the other hand, in reality speakers' use of language relies on more than simple patterns of word co-occurrence – For instance, we use syntactic structures and pragmatic reasoning to supplement the meaning of the individual lexemes we come across (e.g., Fodor, 1995; Grice, 1989 [1975]). It is therefore likely that while LSA captures some of the variability in meaning exhibited by words in context, it does not capture all of it. Indeed, there is a growing body of methods that propose to integrate these two disparate sources of linguistic information (e.g., Pado and Lapata, 2007; Widdows, 2003)

Certainly, the results reported in this paper suggest that enough of the meaning of words and contexts is captured to allow interesting inferences about semantic change and the relatedness of words to be drawn with a reasonable degree of certainty. However, it is possible that some important aspects of meaning are systematically ignored by the analysis. For instance, it remains to be seen whether this method can distinguish between processes like pejoration and amerlioration as they require a fine grained distinction between 'good' and 'bad' meanings.

Regardless of any such limitations, it is clear that important information about meaning can be gathered through a systematic analysis of the contexts in which words appear. Furthermore, phenomena such as the existence of phonaesthemes and the success of LSA in predicting vocabulary acquisition rates, suggest that the acquisition of new vocabulary involves the gleaning of the meaning of words through their context. The role of context in semantic change is therefore likely to be an active one – when a listener encounters a word they are unfamiliar with they are likely to use the context in which it appears, as well as its phonetic composition, as clues to its meaning. Furthermore, if a word is likewise encountered in context in which it is unlikely, this unexpected observation may induce the listener to adjust their representation of both the context and the word in order to increase the overall coherence of the utterance or sentence. As a result, it is possible that examining the contexts in which a word is used in different documents and time periods might be useful not only as a tool for examining the history of a semantic change but also as an instrument for predicting its future progress. Overall, this suggests a dynamic view of the field of semantics – semantics as an ever-changing landscape of meaning. In such a view, semantic change is the norm as the perceived meaning of words keeps shifting to accommodate the contexts in which they are used.

References

Bergen, B. (2004). The Psychological Reality of Phonaesthemes. *Language*, 80(2), 291-311.

Berry, M. W. (1992) *SVDPACK: A Fortran-77 software library for the sparse singular value decomposition.* Tech. Rep. CS-92-159, Knoxville, TN: University of Tennessee.

Berry, M. W., Do, T., O'Brien, G. Vijay, K. Varadh an, S. (1993) *SVDPACKC (Version 1.0) User's Guide*, Tech. Rep. UT-CS-93-194, Knoxville, TN: University of Tennessee.

Bloomfield, L. (1933). *Language*. New York, NY: Holt, Rinehart and Winston.

Campbell, L. (2004) *Historical linguistics: An introduction* 2nd ed. Cambridge, MA: The MIT Press.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990) Indexing by Latent Semantic Analysis. *Journal of the American Society for Information Science*, 41, 391-407.

Ellegård, A. (1953) The Auxiliary Do: the Establishment and Regulation of its Use in English. *Gothenburg Studies in English*, 2. Stockholm: Almqvist and Wiksell.

Firth, J. (1930) *Speech*. London: Oxford University Press.

Firth, J. (1957) *Papers in Linguistics, 1934-1951*, Oxford University Press.

Fodor, J. D. (1995) Comprehending sentence structure. In L. R. Gleitman and M. Liberman, (Eds.), *Invitation to Cognitive Science*, volume 1. MIT Press, Cambridge, MA. 209-246.

Forston, B. W. (2003) An Approach to Semantic Change. In B. D. Joseph and R. D. Janda (Eds.), *The Handbook of Historical Linguistics*. Malden, MA: Blackwell Publishing. 648-666.

Grice, H. P. (1989) [1975]. Logic and Conversation. In *Studies in the Way of Words*. Cambridge, MA: Harvard University Press. 22-40.

Halliday, M. A. K., & Hasan, R. (1976) *Cohesion in English*. London: Longman.

Hock, H. H., and Joseph, B. D. (1996) *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Berlin: Mouton de Gruyter.

Hoey, M. (1991) *Patterns of Lexis in Text*. London: Oxford University Press.

Hutchins, S. S. (1998). The psychological reality, variability, and compositionality of English phonesthemes. *Dissertation Abstracts International*, 59(08), 4500B. (University Microfilms No. AAT 9901857).

Infomap [Computer Software]. (2007). http://infomap-nlp.sourceforge.net/ Stanford, CA.

Kaufmann, S. (2000) Second-order cohesion. *Computational Intelligence.* 16, 511-524.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.

Levin, E., Sharifi, M., & Ball, J. (2006) Evaluation of utility of LSA for word sense discrimination. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, New York City. 77-80.

Marcu, D (2003) Automatic Abstracting, *Encyclopedia of Library and Information Science*, Drake, M. A., ed. 245-256.

Otis K., & Sagi E. (2008) Phonaesthemes: A Corpora-based Analysis. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Pado, S. & Lapata, M. (2007) Dependency-based Construction of Semantic Space Models. *Computational Linguistics*, 33, 161-199.

Riedel E., Dexter S. L., Scharber C., Doering A. (2006) Experimental Evidence on the Effectiveness of Automated Essay Scoring in Teacher Education Cases. *Journal of Educational Computing Research*, 35, 267-287.

Rissanen, M. (1994) The Helsinki Corpus of English Texts. In Kytö, M., Rissanen, M. and Wright S. (eds), *Corpora Across the Centuries: Proceedings of the First International Colloquium on English Diachronic Corpora*. Amsterdam: Rodopi.

Schütze, H. (1996) *Ambiguity in language learning: computational and cognitive models*. CA: Stanford.

Schütze, H. (1998) Automatic word sense discrimination. *Computational Linguistics* 24(1):97-124.

Takayama, Y., Flournoy R., & Kaufmann, S. (1998) *Information Mapping: Concept-Based Information Retrieval Based on Word Associations*. CSLI Tech Report. CA: Stanford.

Takayama, Y., Flournoy, R., Kaufmann, S. & Peters, S. (1999). Information retrieval based on domain-specific word associations. In Cercone, N. and Naruedomkul K. (eds.), *Proceedings of the Pacific Association for Computational Linguistics (PACLING'99)*, Waterloo, Canada. 155-161.

Traugott, E. C. (1999) The Role of Pragmatics in Semantic Change. In J. Verschueren (ed.), *Pragmatics in 1998: Selected Papers from the 6th International Pragmatics Conference, vol. II*. Antwerp: International Pragmatics Association. 93-102.

Traugott, E. C. (2005) Semantic Change. In *Encyclopedia of Language and Linguistics*, 2nd ed., Brown K. ed. Oxford: Elsevier.

Traugott, E. C., and Dasher R. B. (2002) *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Widdows, D. (2003) Unsupervised methods for developing taxonomies by combining syntactic and statistical information. In *Proceedings of the joint Human Language Technology Conference and Annual Meeting of the North American Chapter of the Association for Computational Linguistics*. Edmonton, Canada: Wiemer-Hastings. 197-204.

Widdows, D. (2004) *Geometry and Meaning*. CSLI Publications, CA: Stanford.