Title
Exploring Contextual Influences on Word Meaning via Multiple-Level Similarity Judgments

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Author
Gaylord, Nicholas

Publication Date
2011

Peer reviewed

Exploring Contextual Influences on Word Meaning Via Multiple-Level Similarity Judgments

Nicholas Gaylord (nlgaylord@mail.utexas.edu)
Department of Linguistics, 1 University Station B5100
Austin, TX 78712

Abstract

It is well-known that many words' meanings vary depending on the context in which they are used. This phenomenon has proven difficult to model, and recent work (Klepousniotou, Titone, & Romero, 2008; Erk, McCarthy, & Gaylord, 2009; Brown, 2010) indicates that meaning-in-context varies in a manner much more subtle than would be represented by a dictionary, including reliable fine-grained intuitions about the similarity in meaning of two occurrences of a word in different contexts. This raises the question of how to predict these intuitions of semantic similarity on the basis of contextual information. I present the results of a Magnitude Estimation task in which participants judged the similarity in meaning between pairs of verb occurrences. Stimuli consisted of pairs of sentences containing a present tense intransitive motion verb and a singular definite NP subject (e.g. "The kid runs" vs. "The rabbit runs"). As the only varying component of sentential context, differences between the subject nouns in these pairs are hypothesized to predict intuitions about verbal semantic similarity. I explore four measures of noun similarity – two based on noun animacy, as well as conceptual and distributional similarity measures. I find that individually, all of these measures are significant predictors of verbal semantic similarity judgments, but that the best model of participant judgments combines all four. This is taken as an indication that a proper use of converging sources of evidence enables more accurate, detailed study of the perception of word meaning in context.

Keywords: Magnitude Estimation; semantic similarity; lexical ambiguity; animacy

Introduction

The term *lexical ambiguity* refers to a single word having more than one possible meaning. While the existence of lexical ambiguity is uncontroversial, its precise nature is difficult to pin down, particularly with regard to just how many different meanings a word has, and how that range of possible meanings is structured. One commonly-recognized subtype of lexical ambiguity is *polysemy*, where the different meanings of a word are related to each other, as with "chicken", which can mean (literally) a type of bird, (via metonymy) the meat of that bird, or (metaphorically) a cowardly person. However, the different senses of a word are often not as distinguishable as those of "chicken", making accurate semantic generalizations above the level of individual instances more difficult. It has been argued in the theoretical linguistic literature (Apresjan, 1974) that semantic similarity is actually a matter of degrees, a view that is shared by some lexicographers as well (Hanks, 2000). While there is psycholinguistic evidence for graded similarity in meaning *between* words (Miller & Charles, 1991), the similarity in meaning between different occurrences of a single word is underexplored.

There is growing evidence that word meaning does subtly vary from one context to the next in a manner that may not be readily captured by a dictionary-type approach to meaning (Brown, 2010; Erk et al., 2009). However, the contextual properties that influence the interpretation of a word are not yet clear. This paper presents a preliminary investigation into this question. The goal of this work, broadly put, is to identify what sources of contextual information people use in arriving at a specific interpretation of an ambiguous word. I focus on the identification of broadly relevant properties of sentential context, that can be used to predict intuitions of semantic similarity across a wide range of word occurrences.

Sentential context can be highly complex, including a host of lexical, syntactic, and pragmatic factors. In order to properly begin a detailed study of contextual influences on the perception of word meaning, it is necessary to simplify both the problem domain and the nature of the sentential context itself. The research reported here focuses on the influence of subject nouns on the interpretation of motion verbs, making use of simplified sentences consisting only of an intransitive verb and a subject noun phrase. Stimuli consisted of pairs of sentences containing the same verb, and I obtained semantic similarity judgments for those verb occurrences. While this experiment, on the surface, resembles a sentence comparison task, that is not its primary intention. Rather, it is only through incorporation of contextual information that the meaning of an ambiguous word becomes precise, and this work focuses on comparison of these contextually-induced specific interpretations.

I present the results of a Magnitude Estimation (ME) task (Bard, Robertson, & Sorace, 1996) in which participants were presented with pairs of sentences containing the same verb, and were asked to numerically rate the similarity in meaning of the verb occurrences. These judgments were then analyzed in terms of the similarity of the two subject nouns, which were the only piece of context that varied between the two sentences. Four measures of noun similarity were considered: two based on noun animacy, and two more general similarity measures, one conceptual and one distributional.

Animacy is a linguistic semantic feature with known effects in natural language (Bresnan, 2001; Aissen, 2003; Lee, 2003; Mak, Vonk, & Schreifers, 2006), with the added advantage that it is applicable to all nouns. Any noun is, minimally, either animate or inanimate, but animacy can also be characterized as a scalar feature, in keeping with the hierarchy in (1), from Bresnan (2001):

(1) **Animacy Hierarchy:**
humans > animals > insects > natural forces > plants, inanimate objects > abstract notions

English does not exhibit the same morphosyntactic sensitivity to noun animacy as is seen in other languages, but I follow Comrie (1989) in the assumption that animacy reflects underlying conceptual distinctions, and as such can be applicable as a measure of differences between nouns. As such, I include two measures of noun animacy in this study: a scalar measure as well as a binary living/nonliving distinction reflecting the familiar English animacy distinction.

While animacy has several appealing characteristics as a semantic feature that is useful in characterizing difference between nouns, it is not expected that animacy completely characterizes such differences. I also include two other noun similarity measures in an attempt to account for any similarity not captured by animacy. Distributional similarity ratings for each pair of nouns in the experiment were obtained using Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997) via the online LSA interface from CU Boulder.[1] I also collected human-generated similarity measures for the nouns in the study, as described below. These measures all provide separate, if overlapping, measures of similarity or difference between nouns.

I hypothesized that verbal semantic similarity judgments would be higher for pairs of sentences with more similar subject nouns. As such, participant judgments were hypothesized to be positively correlated with conceptual and distributional noun similarity, and negatively correlated with noun animacy differences. It was further hypothesized that that the four predictors, while not logically independent, would predict participant judgments better collectively than individually. Additionally, noun pairs were hypothesized to influence participant judgments differently depending on verb, with noun difference having a greater effect on judgments for verbs with higher degrees of manner encoding. With the exception of the presence of verb effects, these hypotheses are supported by the data, indicating promise in an effort to predict fine-grained intuitions about word meaning on the basis of contextual properties utilizing converging sources of information.

Background

The research reported here does not purport to be research into the structure of the mental lexicon per se, but rather a study of people's intuitions regarding word meaning in context. The extent to which such intuitions directly reflect the psychological representation of word meanings is left as a question for future study. It is nevertheless the case that perception of a word's meaning in context must depend at least in part upon the nature of one's knowledge of that word. A basic question regarding ambiguous words is whether their range of possible meanings can be adequately represented as a set of discrete entities, such as is commonly found in an dictionary. This is a question that has been taken up in psycholinguistic, theoretical linguistic, and computational linguistic research. With notable exceptions (e.g. Klein & Murphy 2001, 2002), the majority of experimental studies on word meaning do support, for some words at least, a view in which meanings are not discrete and stored separately, but rather at least partially derived during comprehension (Brown, 2010; Frazier & Rayner, 1990; Klepousniotou, 2002; Klepousniotou et al., 2008; Pickering & Frisson, 2001; Williams, 1992). Here, I briefly review these findings as they relate to the study of word meaning based on individual word occurrences.

Broadly speaking, the findings of psycholinguistic studies of polysemy can be divided into two groups: studies that found evidence for a psychological distinction between homonymy and polysemy, and studies that did not. Homonym meanings (e.g. river *bank* vs. financial *bank*) are widely accepted as being distinct and separately stored, but the question is whether the meanings of polysemous words, with their related senses, are structured in the same way. If there is a processing difference between homonyms and polysemous words, this could be due to differences in how their meanings are psychologically represented.

Frazier and Rayner (1990) and Pickering and Frisson (2001) explored homonymy and polysemy via tracking of participants' eye movements and reading times. Both studies found evidence for processing differences between related and unrelated word meanings, concluding that homophonous words force listeners to commit to an interpretation, which can increase comprehension time if the wrong commitment is made. Polysemous words do not force such a commitment. Pickering and Frisson go on to argue for underspecification of lexical entries for polysemous words, in keeping with some theoretical linguistic approaches (e.g. Copestake and Briscoe (1995)). Williams (1992) presents a semantic priming lexical decision task that also supports the homonymy/polysemy distinction, and additionally offers suggestions about the structure of polysemous words' meanings. The results suggest a core meaning from which the various related senses are derived, on the basis of differential priming effects from central and non-central meanings of the word (as with the "soiled" and "obscene" senses of *dirty*, respectively). This is a further indication of the need for detailed study of the processing of ambiguous words in different contexts.

Contrary to the findings discussed above, Klein and Murphy (2001, 2002) fail to find evidence supporting a distinction between homonymy and polysemy, concluding that the senses of a polysemous word are stored separately, despite the intuition that they are related. However, it is possible that Klein and Murphy's results depend partly on their specific stimuli, which contained many polysemous words whose senses were quite distinct, and as such harder to distinguish from homonymy (Klein and Murphy 2001:278, Klepousniotou et al. 2008:1535). Klepousniotou et al. (2008) conducted a study using the same methods as Klein and Murphy (2001), but controlled for the amount of semantic overlap between the senses of polysemous words. Controlling for this, the authors did find a processing difference between high overlap and low-to-moderate overlap words, as such supporting a polysemy-homonymy distinction whereas the original

[1] http://lsa.colorado.edu

study did not. While this does not constitute a complete refutation of Klein and Murphy's findings, it does serve to highlight the fact that relatedness-in-meaning must be conceived of as a matter of degrees.

A graded view of similarity in meaning between word occurrences is implicitly advocated by Cruse (1995, 2000) and similar views are also present within the field of lexicography (Hanks, 2000; Kilgarriff, 2006). Empirical support for this view can be found in Erk et al. (2009) and Brown (2008, 2010). Erk et al. (2009) report on a series of corpus annotation tasks, one of which obtained ordinal semantic similarity ratings for pairs of words in context. They found that annotators did make a range of fine-grained similarity judgments, and that moreover these judgments correlated with graded word sense applicability ratings for those occurrences. Brown (2008, 2010) reports on a semantic priming lexical decision task in which prime-target pairs exhibited varying degrees of semantic similarity. Brown found that response time decreased and response accuracy increased proportional to prime-target semantic similarity. This suggests that there is in fact a processing correlate of our intuitions regarding graded semantic similarity of word occurrences.

Experiment

Adopting the view of graded similarity in meaning between different occurrences of a word, what is needed is an account of from what sources these graded intuitions arise. While the work discussed above generally indicates that word meaning can vary in a fine-grained manner from one context to the next, it does not offer much in the way of explanation regarding what contextual properties drive that variation in meaning. The work reported below represents a first step towards answering this question. Below, I describe the collection of the noun similarity ratings used in analysis of judgments from the primary experiment, which is discussed after.

Collection of Noun Similarity Ratings

This section describes the collection of numeric noun similarity ratings that were subsequently used in the analysis of verbal semantic similarity judgments in the main experiment. Noun similarity measures were obtained via a Magnitude Estimation (ME) task. Participants were presented with pairs of nouns and provided relative ratings of similarity. These ratings were then averaged across participants to obtain a similarity score for each noun pair.

Materials 9 nouns were selected: *adult, kid, cat, rabbit, insect, storm, highway, lane,* and *topic*. These correspond to different levels of the animacy hierarchy in (1), with some levels of the hierarchy being mapped onto by 2 nouns in the set. Additionally, while varying in animacy, these nouns do not vary with respect to other general linguistic semantic features – they are all singular, definite, and genderless. Care was taken to select nouns with similar frequencies, as obtained from Kilgarriff's frequency counts derived from the 100 million word British National Corpus.[2] 9 nouns yield a total of 36 noun pairs, which were rated for similarity by each participant. Within-pair order was varied such that some participants rated the similarity of, e.g. *cat/kid* and others rated the similarity of *kid/cat*.

Procedure The noun similarity rating task was conducted as a paper ME questionnaire. Participants were initially presented with a modulus item (*cake/prize*) that did not contain any of the nouns to be subsequently rated, and asked to assign a number to represent the level of similarity in that pair. They were then instructed, for the remaining 36 noun pairs, to rate their degree of similarity relative to the modulus.

Participants 20 University of Texas undergraduates participated in the task to fulfill a course requirement. Data from 4 participants was discarded due to failure to complete the task.

Results Numerical ratings provided for each noun pair were divided by the modulus value to yield a relative similarity score. Because the same modulus was presented to all participants, scores for each noun pair were then averaged across participants, without collapsing within-pair order, to yield 72 noun pair similarity measures. These judgments of similarity are significantly correlated using Spearman's ρ ($p < .05$) with both animacy measures, but not with distributional similarity measures.

Main Experiment: Collection of Verbal Semantic Similarity Judgments

The main experiment was designed to assess the impact of subject noun properties on participants' perception of verb meaning. In this ME task, participants were presented with pairs of short sentences which had the same verb but different subject nouns (such as *The kid* ***runs*** and *The rabbit* ***runs***), and were asked to rate the similarity in meaning of those verb occurrences. The hypothesis was that verbal semantic similarity judgments will be proportional to conceptual and distributional noun similarity, and inversely proportional to noun animacy distance.

Method

Materials The 9 nouns discussed above were used in creating the stimuli for this experiment. 6 verbs were also selected at different degrees of manner encoding (*move, run, climb, turn, roll* and *wander*). Each of the 36 noun pairs was combined with each verb to yield a total of 216 unique stimuli. These were divided into two groups of 108, each of which contained 3 instances of each noun pair and 18 instances of each verb. Each 108-item group was divided into 3 36-item blocks, each of which contained 1 instance of each noun pair and 6 instances of each verb. Block order was counterbalanced using a 3x3 Latin Square design, and order of sentences within individual stimuli was also counterbalanced, yielding 12 unique versions of the experiment.

[2] http://www.kilgarriff.co.uk/bnc-readme.html

The kid **runs.**
The rabbit **runs.** ___________

Figure 1: An example stimulus from the ME task, containing a pair of sentences and a response blank.

Procedure The verbal semantic similarity rating task was conducted as a paper ME questionnaire. Participants were initially presented with a modulus item (*The runner **stumbles** / the conversation **stumbles***) that did not contain any content words from items to be subsequently rated, and were asked to assign any number that they wished to represent the level of similarity in meaning between the verb occurrences in the modulus. They were then asked to rate the remaining 108 sentence pairs relative to the degree of similarity they perceived in the modulus.

Participants 25 University of Texas undergraduates participated in the experiment. Some participated for course credit, and some were paid $8 for their participation. Data from one participant was discarded due to their indicating that they had not properly understood the task instructions.

Results and Discussion

Each of the 12 unique versions of the questionnaire was completed by two participants. Numerical responses for each item were divided by the modulus score to obtain linear relative similarity scores. Responses for all pairwise combinations of participants that saw the same items were highly significantly correlated ($p<.001$), indicating that participants were performing the task in a reliable and consistent manner. Because responses were non-normally distributed, Spearman's ρ, a nonparametric correlation measure, is used throughout the analysis of this data.

Noun animacy measures were numerically encoded in addition to the numeric conceptual and distributional similarity scores. Nouns were coded for animacy using a 6-point scale, reflecting the 6 levels of the animacy hierarchy in (1). A measure of (dis)similarity between a pair of nouns was obtained via the difference in the nouns' animacy scores, with animacy distance scores ranging from 0 to 5. Nouns were also coded for a binary living/nonliving contrast, which yielded for each stimulus a value of 0, 1, or 2 representing the number of nouns in that stimulus denoting living things. All of these predictors are highly significantly correlated with participant judgments of verbal semantic similarity, as seen in Table 1.

A series of linear mixed-effects models was then fit to the ME data to compare the effectiveness of the various animacy measures and noun similarity measures in accounting for the observed variation in verbal semantic similarity judgments. AIC values (Akaike, 1974) from these models were compared to assess relative goodness of model fit. Models were fit using all possible combinations of the 4 predictors under consideration,[3] with particular interest in models that combined an animacy-derived measure with one of the other two similarity measures. In all cases, models using a combination of predictors fit the ME data better than any model based on a subset of those predictors. This is important to note because of the fact that noun animacy presumably contributes to general conceptual knowledge about nouns. In fact, all combinations of the above 4 predictors, with the surprising exception of human similarity ratings and LSA ratings, are significantly correlated using Spearman's ρ.

Table 1: Correlations between noun similarity measures and verbal semantic similarity judgments

Predictor	ρ	p-value
Animacy Distance	-0.427	2.2e-16
Binary Animacy	0.467	2.2e-16
LSA Noun Similarity	0.178	2.2e-16
Human Noun Similarity	0.365	2.2e-16

The non-independence of the predictors in this model makes it important to demonstrate that they each make an important individual contribution to accounting for the variance in the verbal semantic similarity judgments. Because any model considering multiple predictors had a significantly lower AIC value than any model using a subset of them, it can be seen that each of the four predictors under consideration here is accounting for a different part of the variance in the data. The best-fitting model incorporates all four predictors as both fixed and random effects, without interaction.

Further details on each of these four predictors are contained in Table 2. Because the data from this experiment is non-normally distributed, some assumptions underlying the regression models described above are not supported by the data and it is important to make sure that the models discussed thus far are accurate. In response to this, I used a 3000 iteration bootstrap simulation to generate a large series of models based on the gathered data. In addition to beta values and standard error, which are taken from the linear mixed-effects regression models discussed previously, I report two confidence measures for each predictor: p-values and 95% empirical confidence intervals, both derived from the 3000 simulated model fits. I additionally report mean beta values from the simulation. All of these values are consistent with the ones returned by the parametric models. Figure 2 contains the model beta values for each predictor plotted with error bars representing the empirical 95% confidence intervals in Table 2.

One unexpected observation is the general lack of verb effects in this data. This can be seen in the fact that including Verb as an effect did not substantially improve model fit. It was expected that since the verbs in this study encode a range of degrees of manner, the effects of a noun pair would vary from verb to verb, specifically that interpretations of high

[3]In all cases, predictors, when included in a model, were included as both fixed and random effects.

Table 2: Beta values, standard error (SE), p-values, and empirical confidence intervals (ECI) for best model.

	β_{model}	β_{mean}	SE	p-value	95% ECI (L)	95% ECI (H)
(Intercept)	0.613	0.616	0.143	<.001	0.51	0.721
Animacy Distance	-0.246	-0.247	0.054	<.001	-0.269	-0.226
LSA Noun Similarity	0.606	0.596	0.143	.001	0.23	0.97
Human Noun Similarity	0.624	0.622	0.122	<.001	0.496	0.747
Binary Animacy	0.780	0.782	0.193	<.001	0.724	0.839

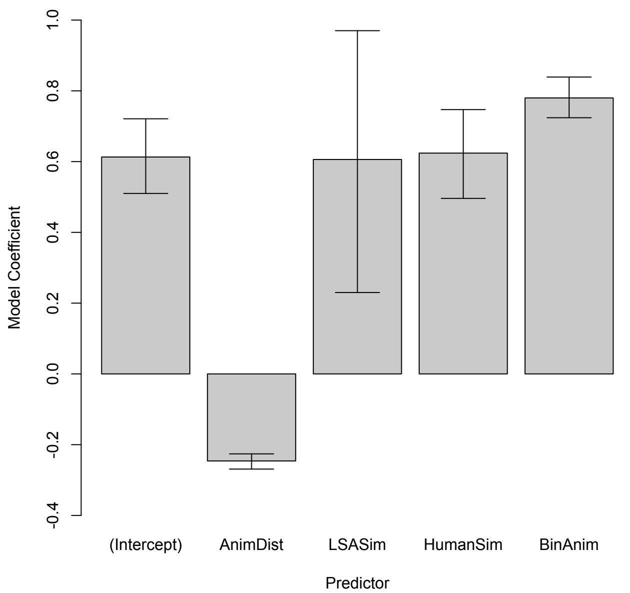

Figure 2: Model estimates and 95% empirical confidence intervals for predictors in best-fitting model.

manner encoding verbs would vary more widely relative to noun pair similarity than low manner encoding verbs. It is not clear from the data that this is the case, though this may be due to the relatively small number of verbs in this study, an issue I return to below in discussing future work.

Another possibility is that participants disregarded the verbs in the stimuli and simply provided noun pair similarity judgments. That, however, does not appear to be the case. The noun similarity ratings, while used to assist in interpreting the verbal judgments, are a long way from fully representing the verbal semantic similarity data. Additionally, performance in the two tasks can be contrasted. In the main experiment, all pairwise correlations of participants that saw the same data were highly significant ($p < 2.2e\text{-}16$), but in the noun similarity rating study, only about half of the pairwise correlations of participants were significant, even at just $p < .05$. This indicates, perhaps somewhat surprisingly, that participants performed more consistently in the verbal semantic similarity task, even though it may appear to be more complex than the noun similarity rating task. One possibility for this is that embedding the verbs in a sentence facilitates comparison of the nouns with respect to alignable differences (Markman & Gentner, 1993; Gentner & Markman, 1997; Gentner & Gunn, 2001) whereas the noun similarity task requires a more abstract comparison. A third possibility is that manner encoding simply does not influence verbal semantic similarity *qua* subject noun similarity. However, the data support no conclusions as yet regarding this question.

General Discussion

I presented the results of a Magnitude Estimation task wherein participants were shown pairs of short sentences with the same verb but different subject nouns. I then showed that participants' judgments of verbal semantic similarity in these pairs were predictable on the basis of four measures of noun similarity: a distributional measure, a conceptual similarity measure, and two measures based on difference in animacy. The results presented here are not a complete account of contextual influences on the meanings of the verbs in these pairs, but they are a first step towards new and better-developed models of contextually-induced word meaning, that are built up from the relative similarity of individual occurrences.

The work reported here is a first step towards accounting for contextual influences on word meaning in a new way. I focused on the comparison of individual word occurrences, and found that quite fine-grained intuitions about word meaning can be accurately predicted on the basis of sentential context. This work does not make any prior assumptions as to the range of possible meanings of a word – rather, it is expected that comparison of a sufficient number of individual occurrences will generate a picture of this range. One primary interest in this work is to begin identifying very general measures of similarity between contexts. Such sources of information could be semantic (such as animacy), conceptual (such as concreteness or the human noun similarity measures discussed here), or distributional (as with the LSA similarity scores). In fact, utilization of a range of converging sources of evidence is expected to be an important strategy in modeling this complex problem.

This work characterizes word meaning in a way that differs from other common characterizations in key respects. For one, it makes no prior assumptions as to the structure of word meaning or the level of granularity at which to characterize it. Additionally, no claims are made as to the separation, or lack thereof, between lexical and conceptual knowledge. While this work does not make specific claims as to the structure of the mental lexicon, the approach to meaning taken here does rely on the availability of general conceptual knowledge in resolving the meaning of a word in context.

The experiment reported here has yielded promising results in the effort to account for contextual effects on people's intuitions of semantic similarity. It is important to discover other contextual factors that, in addition to predicting human intuitions of semantic similarity between a pair of word occurrences, are also independently motivated and sufficiently general to be of relevance across a wide range of occurrences. Exploration of other predictive features will result in better, more informative models of word meaning in context. In addition to exploring other context properties, it is also important to assess the implications of this type of contextually-induced word meaning variation for online sentence comprehension, by way of e.g. response times. Preliminary investigation into this issue is already underway.

Acknowledgments

I would like to thank Colin Bannard and Katrin Erk for their invaluable advice in designing these experiments as well as with analyzing the resultant data.

References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, *21*, 435-483.

Akaike, H. (1974, dec). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Apresjan, J. (1974). Regular polysemy. *Linguistics*, *142*, 5–32.

Bard, E., Robertson, D., & Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, *72*, 32–68.

Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Blackwell Publishers.

Brown, S. (2008). Choosing sense distinctions for WSD: Psycholinguistic evidence. In *Proceedings of ACL-08: HLT, short papers (companion volume)* (p. 249-252). Association for Computational Linguistics.

Brown, S. (2010). *Finding meaning: Sense inventories for improved word sense disambiguation*. Unpublished doctoral dissertation, University of Colorado at Boulder.

Comrie, B. (1989). *Language universals and linguistic typology*. University of Chicago Press.

Copestake, A., & Briscoe, T. (1995). Semi-productive polysemy and sense extension. *Journal of Semantics*, *12*, 15–67.

Cruse, D. (1995). Polysemy and related phenomena from a cognitive linguistic viewpoint. In P. Saint-Dizier & E. Vegas (Eds.), *Computational lexical semantics* (p. 33-49). Cambridge: Cambridge University Press.

Cruse, D. (2000). Aspects of the micro-structure of word meanings. In Y. Ravin & C. Leacock (Eds.), *Polysemy: Theoretical and computational approaches* (p. 30-51). New York: Oxford University Press.

Erk, K., McCarthy, D., & Gaylord, N. (2009). Investigations on word senses and word usages. In *Proceedings of acl 2009*.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitements: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*, 181–200.

Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory and Cognition*, *21*, 565–577.

Gentner, D., & Markman, A. (1997). Structural alignment in analogy and similarity. *American Psychologist*, *52*, 45–56.

Hanks, P. (2000). Do word meanings exist? *Computers and the Humanities*, *34*, 205-215.

Kilgarriff, A. (2006). Word senses. In E. Agirre & P. Edmonds (Eds.), *Word sense disambiguation: Algorithms and applications* (pp. 29–46). Dordrecht: Springer.

Klein, D., & Murphy, G. (2001). The representation of polysemous words. *Journal of Memory and Language*, *45*, 259-282.

Klein, D., & Murphy, G. (2002). Paper has been my ruin: conceptual relations of polysemous senses. *Journal of Memory and Language*, *47*, 548-570.

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, *81*, 205–223.

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1534–1543.

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

Lee, H. (2003). Prominence mismatch and markedness reduction in word order. *Natural Language and Linguistic Theory*, 617-680.

Mak, W., Vonk, W., & Schreifers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, *54*, 466–490.

Markman, A., & Gentner, D. (1993). Splitting the difference: A structural alignment view of similarity. *Journal of Memory and Language*, *32*, 517–535.

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*(1), 1–28.

Pickering, M., & Frisson, S. (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 556–573.

Williams, J. (1992). Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research*, *21*, 193–218.