

# What Is Word Meaning, Really? (And How Can Distributional Models Help Us Describe It?)

Katrin Erk

Department of Linguistics  
University of Texas at Austin  
katrin.erk@mail.utexas.edu

## Abstract

In this paper, we argue in favor of re-considering models for word meaning, using as a basis results from cognitive science on human concept representation. More specifically, we argue for a more flexible representation of word meaning than the assignment of a single best-fitting dictionary sense to each occurrence: Either use dictionary senses, but view them as having fuzzy boundaries, and assume that an occurrence can activate multiple senses to different degrees. Or move away from dictionary senses completely, and only model similarities between individual word usages. We argue that distributional models provide a flexible framework for experimenting with alternative models of word meanings, and discuss example models.

## 1 Introduction

Word sense disambiguation (WSD) is one of the oldest problems in computational linguistics (Weaver, 1949) and still remains challenging today. State-of-the-art performance on WSD for WordNet senses is at only around 70-80% accuracy (Edmonds and Cotton, 2001; Mihalcea et al., 2004). The use of coarse-grained sense groups (Palmer et al., 2007) has led to considerable advances in WSD performance, with accuracies of around 90% (Pradhan et al., 2007). But this figure averages over lemmas, and the problem remains that while WSD works well for some lemmas, others continue to be tough.

In WSD, polysemy is typically modeled through a list of dictionary senses thought to be mutually disjoint, such that each occurrence of a word is characterized through one best-fitting dictionary sense. Accordingly, WSD is typically

framed as a classification task. Interestingly, the task of assigning a single best word sense is very hard for human annotators, not just machines (Kilgarriff and Rosenzweig, 2000).

In this paper we advocate the exploration of alternative computational models of word meaning. After all, one possible reason for the continuing difficulty of (manual as well as automatic) word sense assignment is that the prevailing model might be suboptimal. We explore three main hypotheses. The first builds on research on the human concept representation that has shown that concepts in the human mind do not work like sets with clear-cut boundaries; they show graded membership, and there are typical members as well as borderline cases (Rosch, 1975; Hampton, 2007). Accordingly, **(A)** we will suggest that **word meaning may be better modeled using a graded notion of sense membership** than through concepts with hard boundaries. Second, even if senses have soft boundaries, the question remains of whether they are disjoint. **(B)** We will argue in favor of a framework where **multiple senses may apply to a single occurrence, to different degrees**. This can be viewed as a dynamical grouping of senses for each occurrence, in contrast to static sense groups as in Palmer et al. (2007). The first two hypotheses still rely on an existing sense list. However, there is no universal agreement across dictionaries and across tasks on the number of senses that words have (Hanks, 2000). Kilgarriff (1997) even argues that general, task-independent word senses do not exist. **(C)** **By focusing on individual occurrences (usages) of a lemma and their degree of similarity, we can model word meaning without recourse to dictionary senses**.

In this paper, we are going to argue in favor of the use of vector space as a basis for alternative models of word meaning. Vector space models have been used widely to model word sense (Lund

and Burgess, 1996; Deerwester et al., 1990; Landauer and Dumais, 1997; Sahlgren and Karlgren, 2005; Padó and Lapata, 2007), their central property being that proximity in space can be used to predict semantic similarity. By viewing word occurrences as points in vector space, we can model word meaning without recourse to senses. An additional advantage of vector space models is that they are also widely used in human concept representation models, yielding many modeling ideas that can be exploited for computational models.

In Section 2 we review the evidence that word sense is a tough phenomenon to model, and we lay out findings that support hypotheses (A)-(C). Section 4 considers distributional models that represent word meaning without recourse to dictionary senses, following (C). In Section 5 we discuss possibilities for embedding dictionary senses in vector space in a way that respects points (A) and (B).

## 2 Computational and cognitive models of word meaning

In this section, we review the problems of (manual and automatic) sense assignment, and we discuss cognitive models of concept representation and polysemy, following the three hypotheses laid out in the introduction.

**Word sense assignment.** In computational linguistics, the problem of polysemy is typically phrased as one of choosing one best-fitting sense for the given occurrence out of a dictionary-defined sense list. However, this is a hard task both for humans and for machines. With WordNet (Fellbaum, 1998), the electronic lexicon resource that is currently most widely used in computational linguistics, inter-annotator agreement (ITA) lies in the range of 67% to 78% (Landes et al., 1998; Snyder and Palmer, 2004; Mihalcea et al., 2004), and state-of-the-art WSD systems achieve accuracy scores of 73% to 77% (Edmonds and Cotton, 2001; Mihalcea et al., 2004). This problem is not specific to WordNet: Analyses with the HECTOR dictionary led to similar numbers (Kilgarriff and Rosenzweig, 2000). Sense granularity has been suggested as a reason for the difficulty of the task (Palmer et al., 2007). And in fact, the use of more coarse-grained senses leads to greatly ITA as well as WSD accuracy, with about a 10% improvement for either measure (Palmer et al., 2007; Pradhan et al., 2007). In OntoNotes (Hovy et al., 2006), an ITA of 90% is

used as the criterion for the construction of coarse-grained sense distinctions. However, intriguingly, for some high-frequency lemmas such as *leave* this ITA threshold is not reached even after multiple re-partitionings of the semantic space (Chen and Palmer, 2009) – indicating that the meaning of these words may not be separable into senses distinct enough for consistent annotation. A recent analysis of factors influencing ITA differences between lemmas (Passonneau et al., 2010) found three main factors: sense concreteness, specificity of the context in which a target word occurs, and similarity between senses. It is interesting to note that only one of those factors, the third, can be addressed through a change of dictionary.

More radical solutions than sense grouping that have been proposed are to restrict the task to determining predominant sense in a given domain (McCarthy et al., 2004), or to work directly with paraphrases (McCarthy and Navigli, 2009).

**(A) Graded sense membership.** Research on the human concept representation (Murphy, 2002; Hampton, 2007) shows that categories in the human mind are not simply sets with clear-cut boundaries. Some items are perceived as more typical than others (Rosch, 1975; Rosch and Mervis, 1975). Also, some items are clear members, others are rated as borderline (Hampton, 1979). On borderline items, people are more likely to change their mind about category membership (McCloskey and Glucksberg, 1978). However, these results concern mental concepts, which raises the question of the relation between mental concepts and word senses. This relation is discussed in most depth by Murphy (1991; 2002), who argues that while not every human concept is associated with a word, word meanings show many of the same phenomena as concepts in general; word meaning is “made up of pieces of conceptual structure”. In cognitive linguistics there has been much work on word meaning based on models with graded membership and typically effects (Coleman and Kay, 1981; Lakoff, 1987; Cruse, 1986; Taylor, 1989).

**(B) Multiple senses per occurrence.** While most manual word sense annotation efforts allow annotators to assign more than one dictionary sense to an occurrence, this is typically phrased as an exception rather than the default. In the recent WSSim annotation study (Erk et al., 2009),

Sentence	Senses							Annotator
	1	2	3	4	5	6	7	
This question provoked <b>arguments</b> in America about the Norton Anthology of Literature by Women, some of the contents of which were said to have had little value as literature.	1	4	4	2	1	1	3	Ann. 1
	4	5	4	2	1	1	4	Ann. 2
	1	4	5	1	1	1	1	Ann. 3

Table 1: From (Erk et al., 2009): A sample annotation from the WSsim dataset. The senses are: 1:statement, 2:controversy, 3:debate, 4:literary argument, 5:parameter, 6:variable, 7:line of reasoning

we asked three human annotators to judge the applicability of WordNet senses on a graded scale of 1 (*completely different*) to 5 (*identical*) and giving a rating for *each* sense rather than picking one. Table 1 shows an example sentence with annotator ratings for the senses of the target *argument*. For this sentence, the annotators agree that senses 2 and 3 are highly applicable, but there also individual differences in the perceived meaning: Only annotator 2 views sense 1 as applying to a high degree. In an annotation setting with graded judgments, it does not make sense to measure exact agreement on judgments. We instead evaluated ITA using Spearman’s rho, a nonparametric correlation test, finding highly significant correlations ( $p \ll 0.001$ ) between each pair of annotators, as well as highly significant correlations with the results of a previous, traditional word sense annotation of the same dataset. The annotators made use of the complete scale (1-5), often opting for intermediate values of sense applicability. In addition, we tested whether there were groups of senses that always got the same ratings on any given sentence (which would mean that the annotators implicitly used more coarse-grained senses). What we found instead is that the annotators seemed to have mixed and matched senses for the individual occurrences in a dynamic fashion.

**(C) Describing word meaning without dictionary senses.** In lexicography, Kilgarriff (1997) and Hanks (2000) cast doubt on the existence of task-independent, distinct senses. In cognitive science, Kintsch (2007) calls word meaning “fluid and flexible”. And some researchers in lexical semantics have suggested that word meanings lie on a continuum between clear cut cases of ambiguity on the one hand, and on the other hand vagueness where clear cut boundaries do not hold (Tuggy, 1993). There are some psychological studies on whether different senses of a polysemous word are represented separately in the mind or whether there is some joint representation. However, so far the evidence is inconclusive

1) *We study the methods and concepts that each writer uses to defend the cogency of legal, deliberative, or more generally political prudence against explicit or implicit charges that practical thinking is merely a knack or form of cleverness.*

2) *Eleven CIRA members have been convicted of criminal charges and others are awaiting trial.*

Figure 1: From (Erk et al., 2009): A sense pair from the USim dataset, for the target *charge.n*. Annotator judgments: 2,3,4

and varies strongly with the experimental setting. Some studies found evidence for a separate representation (Klein and Murphy, 2001; Pylkkanen et al., 2006). Brown (2008) finds a linear change in semantic similarity effects with sense distance, which could possibly point to a continuous representation of word meaning without clear sense boundaries. But while there is no definitive answer yet on the question of the mental representation of polysemy, a computational model that does not rely on distinct senses has the advantage of making fewer assumptions. It also avoids the tough lexicographic problem mentioned above, of deciding on a best set of senses for a given domain.

In the recent USim annotation study (Erk et al., 2009), we tested whether human annotators could reliably and consistently provide word meaning judgments without the use of dictionary senses. Three annotators rated the similarity of pairs of occurrences (usages) of a common target word, again on a scale of 1-5. Figure 1 shows an example, with the corresponding annotator judgments. The results on this task were encouraging: Again using correlation to measure ITA, we found a highly significant correlation ( $p \ll 0.001$ ) between the judgments of each pair of annotators. Furthermore, there was a strong correlation on judgments given with and without the use of dictionary senses (USim versus WSsim) for the same data.

### 3 Vector space models of word meaning in isolation

This section gives a brief overview of the use of vector spaces to model concepts and word meaning in cognition and computational linguistics.

In two of the current main theories of concept representation, feature vectors play a prominent role. Prototype theory (Hampton, 1979; Smith and Medin, 1981) models degree of category membership through similarity to a single prototype. Exemplar models (Medin and Schaffer, 1978; Nosofsky, 1992; Nosofsky and Palmeri, 1997) represent a concept as a collection of all previously seen exemplars and compute degree of category membership as similarity to stored exemplars. Both prototypes and exemplars are typically represented as feature vectors. Many models represent a concept as a region rather than a point in space, often characterized by a feature vector plus a separate dimension weight vector (Smith et al., 1988; Hampton, 1991; Gärdenfors, 2004). The features are individually meaningful and interpretable and include sensory and motor features as well as function and taxonomic features. There are several datasets with features elicited from human subjects (McRae et al., 2005; Vigliocco et al., 2004).

In computational linguistics, distributional models represent the meaning of a word as a vector in a high-dimensional space whose dimensions characterize the contexts in which the word typically occurs (Lund and Burgess, 1996; Landauer and Dumais, 1997; Sahlgren and Karlgren, 2005; Padó and Lapata, 2007). In the simplest case, the dimensions are context words, and the values are co-occurrence counts. In contrast to spaces used in cognitive science, the dimensions in distributional models are typically not interpretable (though see Almuhareb and Poesio (2005), Baroni et al. (2010)). A central property of distributional models is that proximity in vector space is a predictor of semantic similarity. These models have been used successfully in NLP (Deerwester et al., 1990; Manning et al., 2008), as well as in psychology (Landauer and Dumais, 1997; Lowe and McDonald, 2000; McDonald and Ramscar, 2001).

### 4 Vector space models of word meaning in context

If we want to represent word meaning through individual usages and their similarity only, without the use of dictionary senses (along hypothesis

(C)), distributional models are an obvious choice, if we can just represent each individual usage as a point in space. However, vector space models have mostly been used to represent the meaning of a word in isolation: The vector for a word is computed by summing over all its corpus occurrences, thereby summing over all its meanings. There are a few vector space models of meaning *in context*, though they differ in what it is that they model. One group of models computes a single vector for a whole sentence, encoding both the words and the syntactic structure (Smolensky, 1990; B. Coecke and Clark, 2010). In this case, the dimensionality of the vectors varies with the syntactic complexity of the sentence in question. A second group also computes a single vector for a whole expression, but the vector for a larger expression is a combination of the word vectors for the words occurring in the expression (Landauer and Dumais, 1997; Mitchell and Lapata, 2008). Syntactic structure is not encoded. The resulting vector, of the same dimensionality as the word vectors, is then a combination of the contexts in which the words of the sentence occur. A third group of approaches derives a separate vector for each word in a given sentence (Erk and Padó, 2008; Thater et al., 2009; Erk and Padó, 2010). While an approach of the second type would derive a single, joint vector for, say, the expression *catch a ball*, an approach from the third group would derive two vectors, one for the word *catch* in the context of *ball*, and one for the word *ball* in the context of *catch*. In this third group, the dimensionality of a vector for a word in context is the same as for a word in isolation.

In this paper, we focus on the third type of approaches. Our aim is to study alternatives to dictionary senses for characterizing word meaning. So we need a meaning characterization for each individual word in a given sentence context, rather than a single vector for a larger expression.

We can also classify distributional approaches to word meaning in context into *prototype-* and *exemplar-based* approaches. Prototype-based approaches first compute a (prototype) vector for each word in isolation, then modify this vector according to the context in a given occurrence (Landauer and Dumais, 1997; Mitchell and Lapata, 2008; Erk and Padó, 2008; Thater et al., 2009). Typical methods for combining prototype vectors are addition, component-wise multiplication (introduced by Mitchell and Lap-

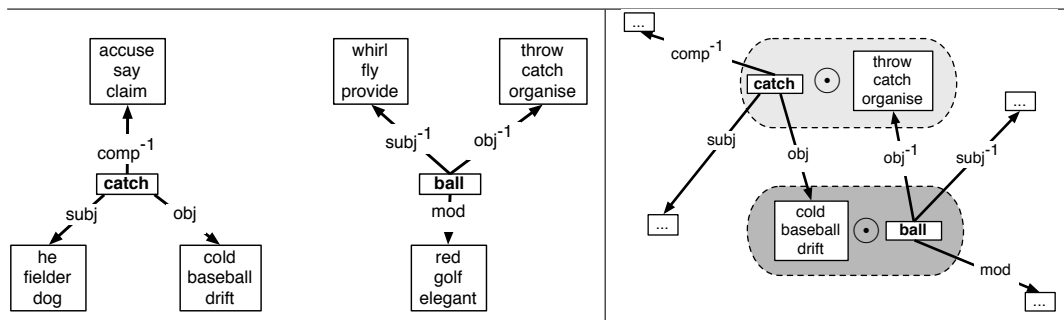


Figure 2: From (Erk and Padó, 2008): Left: Vector representations for verb *catch* and noun *ball*. Lexical information plus selectional preferences. Right: Computing context-specific meaning by combining predicate and argument via selectional preference vectors

ata (2008)), and component-wise minimum. Then there are *multiple prototype* approaches that statically cluster synonyms or occurrences to induce word senses (Schütze, 1998; Pantel and Lin, 2002; Reisinger and Mooney, 2010). *Exemplar-based* approaches represent a word in isolation as a collection of its occurrences or paraphrases, then select only the contextually appropriate exemplars for a given occurrence context (Kintsch, 2001; Erk and Padó, 2010). In this paper we focus on the first and third group of approaches, as they do not rely on knowledge of how many word senses (clusters) there should be.

**A structured vector space model for word meaning in context.** In Erk and Padó (2008), we proposed the *structured vector space model (SVS)*, which relies solely on syntactic context for computing a context-specific vector. It is a prototype-based model, and called *structured* because it explicitly represents argument structure, using multiple vectors to represent each word. Figure 2 (left) illustrates the representation. A word, for example *catch*, has one vector describing the meaning of the word itself, the *lexical vector*  $\vec{catch}$ . It is a vector for the word in isolation, as is usual for prototype-based models. In addition, the representation for *catch* contains further vectors describing the *selectional preferences* for each argument position. The *obj* preference vector of *catch* is computed from the lexical vectors of all words that have been observed as direct objects of *catch* in some syntactically parsed corpus. In the example in Figure 2, we have observed the direct objects *cold*, *baseball*, and *drift*. In the simplest case, the *obj* preference vector of *catch* is then computed as the (weighted) sum of the three vectors  $\vec{cold}$ ,  $\vec{baseball}$  and  $\vec{drift}$ . Likewise, *ball* is represented by one vector for *ball* itself, one for *ball*'s

preferences for its modifiers (*mod*), one vector for the verbs of which it is a subject ( $\vec{subj}^{-1}$ ), and one for the verbs of which is an object ( $\vec{obj}^{-1}$ ).

The vector for *catch* in a given context, say in the context *catch ball*, is then computed as illustrated on the right side of Figure 2: The lexical vector  $\vec{catch}$  is combined with the  $\vec{obj}^{-1}$  vector of *ball*, modifying the vector *catch* in the direction of verbs that typically take *ball* as an object. For the vector combination, any of the usual operations can be used: addition, component-wise multiplication, or minimum. Likewise, the lexical vector  $\vec{ball}$  is combined with the *obj* preference vector of *catch* to compute the meaning of *ball* in the context *catch ball*.

The standard evaluation for vector models of meaning in context is to predict paraphrase appropriateness. Paraphrases always apply to a word meaning, not a word. For example, *contract* is an appropriate paraphrase for *catch* in the context *John caught the flu*, but it is not an appropriate paraphrase in the context *John caught a butterfly*. A vector space model can predict paraphrase appropriateness as the similarity (measured, for example, using Cosine) of the context-specific vector of *catch* with the lexical vector of *contract*: The more similar the vectors, the higher the predicted appropriateness of the paraphrase. We evaluated SVS on two datasets. The first is a tightly controlled psycholinguistic dataset of subject/verb pairs with paraphrases for the verbs only (Mitchell and Lapata, 2008). The other is the Lexical Substitution dataset, which has annotator-generated paraphrases for target words in a larger sentential context and which is thus closer to typical NLP application scenarios (McCarthy and Navigli, 2009). SVS showed comparable performance to the model by Mitchell and Lapata (2008) on the

former dataset, and outperformed the Mitchell and Lapata model on the latter.

One obvious extension is to use all available syntactic context, instead of focusing on a single syntactic neighbor. We found no improvement on SVS in a straightforward extension to additional syntactic context items (Erk and Padó, 2009). However, Thater et al. (2009) did achieve better performance with a different model that used all syntactic context.

**Taking larger context into account in an exemplar-based model.** But even if we take the complete local syntactic context into account, we are missing some evidence, in particular non-local information. The word *ball* is interpreted differently in sentences (1a) and (1b)<sup>1</sup> even though its predicate *ran* has more or less the same meaning in both sentences. What is different is the subject of *ran*, *player* versus *debutante*, which is not a direct syntactic neighbor of the ambiguous word *ball*.

- (1) (a) the player ran to the ball  
(b) the debutante ran to the ball

Even though we are not using dictionary senses, the types of evidence that should be useful for computing occurrence-specific vectors should be the same as for traditional WSD; and one of the main type of features used there is bag-of-words context. In (Erk and Padó, 2010), we proposed an exemplar-based model of word meaning in context that relied on bag-of-words context information from the whole sentence, but did not use syntactic information. The model assumes that each target lemma is represented by a set of exemplars, where an exemplar is a sentence in which the target lemma occurs. Polysemy is then modeled by *activating* (selecting) relevant exemplars of a target lemma in a given occurrence *s*.<sup>2</sup> Both the exemplars and the occurrence *s* are modeled as vectors. We simply use first-order vectors that reflect the number of times each word occurs in a given sentence. The activated exemplars are then simply the ones whose vectors are most similar to the vector of *s*. The results that we achieved with the exemplar-based model on the Lexical Substitution dataset were considerably better than

<sup>1</sup>These two examples are due to Ray Mooney.

<sup>2</sup>Instead of the binary selection of each exemplar that this model uses, it would also be possible to assign each exemplar a weight, making it partially selected.

those achieved with any of the syntax-based approaches (Erk and Padó, 2008; Erk and Padó, 2009; Thater et al., 2009).

While prototype models compute a vector by first summing over all observed occurrences and then having to suppress dimensions that are not contextually appropriate, exemplar models only take contextually appropriate exemplars into account in the first place, which is conceptually simpler and thus more attractive. But there are still many open questions, in particular the best combination of bag-of-words context and syntactic context as evidence for computing occurrence-specific vector representations.

## 5 The role of dictionary senses

Word meaning models that rely only on individual word usages and their similarities are more flexible than dictionary-based models and make less assumptions. On the other hand, dictionaries offer not just sense lists but also a wealth of information that can be used for inferences. WordNet (Fellbaum, 1998) has relations between words and between synsets, most importantly synonymy and hyponymy. VerbNet (Kipper et al., 2000) specifies semantic properties of a predicate’s arguments, as well as relations between the arguments.

In this section we discuss approaches for embedding dictionary senses in a distributional model in a way that supports hypotheses (A) and (B) (graded sense membership, and description of an occurrence through multiple senses) and that supports testing the applicability of dictionary-based inference rules.

**Mapping dictionary senses to points in vector space.** Dictionary senses can be mapped to points in vector space very straightforwardly if we have sense-annotated corpus data. In that case, we can compute a (prototype) vector for a sense from all corpus occurrences annotated with that sense. We used this simple model (Erk and McCarthy, 2009) to predict the graded sense applicability judgments from the WSim dataset. (See Section 2 for more information on this dataset.) The predictions of the vector space model significantly correlate with annotator judgments. In comparison with an approach that uses the confidence levels of a standard WSD model as predictions, the vector space model shows higher recall but lower precision – for definitions of precision and recall that are adapted to the graded case.

Another way of putting the findings is to say that the WSD confidence levels tend to under-estimate sense applicability, while the vector space model tends to over-estimate it.

**Attachment sites for inference rules.** As discussed above, vector space models for word meaning in context are typically evaluated on paraphrase applicability tasks (Mitchell and Lapata, 2008; Erk and Padó, 2008; Erk and Padó, 2009; Thater et al., 2009). They predict the applicability of a paraphrase like (2) based on the similarity between a context-specific vector for the lemma (here, *catch*) and a context-independent vector for the paraphrase. (in this case, *contract*).

$$X \text{ catch } Y \rightarrow X \text{ contract } Y \quad (2)$$

Another way of looking at this is to consider the inference rule (2) to be *attached* to a point in space, namely the vector for *contract*, and to trigger the inference rule for an occurrence of *catch* if it is close enough to the *attachment site*. If we know the WordNet sense of *contract* for which rule (2) holds – it happens to be sense 4 –, we can attach the rule to a vector for sense 4 of *contract*, rather than a vector computed from all occurrences of the lemma. Note that when we use dictionaries as a source for inference rules, for example by creating an inference rule like (2) for each two words that share a synset and for each direct hyponym/hypernym pair, we do know the WordNet sense to which each inference rule attaches.

**Mapping dictionary senses to regions in vector space.** In Erk (2009) we expand on the idea of tying inference rules to attachment sites by representing a word sense not as a point but as a *region* in vector space. The extent of the regions is estimated through the use of both positive exemplars (occurrences of the word sense in question), and negative exemplars (occurrences of other words). The computational models we use are inspired by cognitive models of concept representation that represent concepts as regions (Smith et al., 1988; Hampton, 1991), in particular adopting Shepard’s law (Shepard, 1987), which states that perceived similarity to an exemplar decreases *exponentially* with distance from its vector.

In the longer term, the goal for the association of inference rules with attachment sites is to obtain a principled framework for reasoning with partially applicable inference rules in vector space.

## 6 Conclusion and outlook

In this paper, we have argued that it may be time to consider alternative computational models of word meaning, given that word sense disambiguation, after all this time, is still a tough problem for humans as well as machines. We have followed three hypotheses. The first two involve dictionary senses, suggesting that (A) senses may best be viewed as applying to a certain degree, rather than in a binary fashion, and (B) that it may make sense to describe an occurrence through multiple senses as a default rather than an exception. The third hypothesis then departs from dictionary senses, suggesting (C) focusing on individual word usages and their similarities instead. We have argued that distributional models are a good match for word meaning models following hypotheses (A)-(C): They can represent individual word usages as points in vector space, and they can also represent dictionary senses in a way that allows for graded membership and overlapping senses, and we have discussed some existing models, both prototype-based and exemplar-based.

One big question is, of course, about the usability of these alternative models of word meaning in NLP applications. Will they do better than dictionary-based models? The current evaluations, testing paraphrase applicability in context, are a step in the right direction, but more task-oriented evaluation schemes have to follow.

We have argued that it makes sense to look to cognitive models of mental concept representation. They are often based on feature vectors, and there are many interesting ideas in these models that have not yet been used (much) in computational models of word meaning. One of the most exciting ones, perhaps, is that cognitive models often have interpretable dimensions. While dimensions of distributional models are usually not individually interpretable, there are some first models (Almuhareb and Poesio, 2005; Baroni et al., 2010) that use patterns to extract meaningful dimensions from corpus data. This offers many new perspectives: For which tasks can we improve performance by selecting dimensions that are meaningful specifically for that task (as in Mitchell et al. (2008))? Can interpretable dimensions be used for inferences? And, when we are computing vector space representations for word meaning *in context*, is it possible to select meaningful dimensions that are appropriate for a given context?

**Acknowledgements.** This work was supported in part by National Science Foundation grant IIS-0845925, and by a Morris Memorial Grant from the New York Community Trust.

## References

- A. Almuhareb and M. Poesio. 2005. Finding concept attributes in the web. In *Proceedings of the Corpus Linguistics Conference*, Birmingham.
- M. Sadrzadeh B. Coecke and S. Clark. 2010. Mathematical foundations for a compositional distributed model of meaning. *Lambek Festschrift, Linguistic Analysis*, 36.
- M. Baroni, B. Murphy, E. Barbu, and M. Poesio. 2010. Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science*, 34(2):222–254.
- S. W. Brown. 2008. Choosing sense distinctions for WSD: Psycholinguistic evidence. In *Proceedings of ACL/HLT*, Columbus, OH.
- J. Chen and M. Palmer. 2009. Improving English verb sense disambiguation performance with linguistically motivated features and clear sense distinction boundaries. *Journal of Language Resources and Evaluation, Special Issue on SemEval-2007*, 43:181–208.
- L. Coleman and P. Kay. 1981. The English word “lie”. *Linguistics*, 57.
- D. A. Cruse. 1986. *Lexical Semantics*. Cambridge University Press.
- S. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnaas, and R. A. Harshman. 1990. Indexing by latent semantic analysis. *Journal of the Society for Information Science*, 41(6):391–407.
- P. Edmonds and S. Cotton, editors. 2001. *Proceedings of the SensEval-2 Workshop*, Toulouse, France. ACL. See <http://www.sle.sharp.co.uk/senseval>.
- K. Erk and D. McCarthy. 2009. Graded word sense assignment. In *Proceedings of EMNLP*, Singapore.
- K. Erk and S. Padó. 2008. A structured vector space model for word meaning in context. In *Proceedings of EMNLP*, Honolulu, HI.
- K. Erk and S. Padó. 2009. Paraphrase assessment in structured vector space: Exploring parameters and datasets. In *Proceedings of the EACL Workshop on Geometrical Methods for Natural Language Semantics (GEMS)*.
- K. Erk and S. Padó. 2010. Exemplar-based models for word meaning in context. In *Proceedings of the ACL*, Uppsala.
- K. Erk, D. McCarthy, and N. Gaylord. 2009. Investigations on word senses and word usages. In *Proceedings of ACL*, Singapore.
- Katrin Erk. 2009. Representing words as regions in vector space. In *Proceedings of CoNLL*.
- C. Fellbaum, editor. 1998. *WordNet: An electronic lexical database*. MIT Press, Cambridge, MA.
- P. Gärdenfors. 2004. *Conceptual spaces*. MIT press, Cambridge, MA.
- J. A. Hampton. 1979. Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 18:441–461.
- J. A. Hampton. 1991. The combination of prototype concepts. In P. Schwanenflugel, editor, *The psychology of word meanings*. Lawrence Erlbaum Associates.
- J. A. Hampton. 2007. Typicality, graded membership, and vagueness. *Cognitive Science*, 31:355–384.
- P. Hanks. 2000. Do word meanings exist? *Computers and the Humanities*, 34(1-2):205–215(11).
- E. H. Hovy, M. Marcus, M. Palmer, S. Pradhan, L. Ramshaw, and R. Weischedel. 2006. OntoNotes: The 90% solution. In *Proceedings of HLT-NAACL*, pages 57–60, New York.
- A. Kilgarriff and J. Rosenzweig. 2000. Framework and results for English Senseval. *Computers and the Humanities*, 34(1-2):15–48.
- A. Kilgarriff. 1997. I don’t believe in word senses. *Computers and the Humanities*, 31(2):91–113.
- W. Kintsch. 2001. Predication. *Cognitive Science*, 25:173–202.
- W. Kintsch. 2007. Meaning in context. In T.K. Landauer, D. McNamara, S. Dennis, and W. Kintsch, editors, *Handbook of Latent Semantic Analysis*, pages 89–105. Erlbaum, Mahwah, NJ.
- K. Kipper, H.T. Dang, and M. Palmer. 2000. Class-based construction of a verb lexicon. In *Proceedings of AAAI/IAAI*.
- D.E. Klein and G.L. Murphy. 2001. The representation of polysemous words. *Journal of Memory and Language*, 45:259–282.
- G. Lakoff. 1987. *Women, fire, and dangerous things*. The University of Chicago Press.
- T. Landauer and S. Dumais. 1997. A solution to Platos problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240.
- S. Landes, C. Leacock, and R. Teng. 1998. Building semantic concordances. In C. Fellbaum, editor, *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA.



- W. Lowe and S. McDonald. 2000. The direct route: Mediated priming in semantic space. In *Proceedings of the Cognitive Science Society*, pages 675–680.
- K. Lund and C. Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, 28:203–208.
- C. D. Manning, P. Raghavan, and H. Schütze. 2008. *Introduction to Information Retrieval*. Cambridge University Press.
- D. McCarthy and R. Navigli. 2009. The English lexical substitution task. *Language Resources and Evaluation*, 43(2):139–159. Special Issue on Computational Semantic Analysis of Language: SemEval-2007 and Beyond.
- D. McCarthy, R. Koeling, J. Weeds, and J. Carroll. 2004. Finding predominant senses in untagged text. In *Proceedings of ACL*, Barcelona.
- M. McCloskey and S. Glucksberg. 1978. Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, 6:462–472.
- S. McDonald and M. Ramscar. 2001. Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. In *Proceedings of the Cognitive Science Society*, pages 611–616.
- K. McRae, G. S. Cree, M. S. Seidenberg, and C. McNorgan. 2005. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4):547–559.
- D. L. Medin and M. M. Schaffer. 1978. Context theory of classification learning. *Psychological Review*, 85:207–238.
- R. Mihalcea, T. Chklovski, and A. Kilgariff. 2004. The Senseval-3 English lexical sample task. In *Proceedings of SensEval-3*, Barcelona.
- J. Mitchell and M. Lapata. 2008. Vector-based models of semantic composition. In *Proceedings of ACL*, Columbus, OH.
- T. Mitchell, S. Shinkareva, A. Carlson, K. Chang, V. Malave, R. Mason, and M. Just. 2008. Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880):1191–1195.
- G. L. Murphy. 1991. Meaning and concepts. In P. Schwanenflugel, editor, *The psychology of word meanings*. Lawrence Erlbaum Associates.
- G. L. Murphy. 2002. *The Big Book of Concepts*. MIT Press.
- R. M. Nosofsky and T. J. Palmeri. 1997. An exemplar-based random walk model of speeded classification. *Psychological Review*, 104(2):266–300.
- R. M. Nosofsky. 1992. Exemplars, prototypes, and similarity rules. In A. Healy, S. Kosslyn, and R. Shiffrin, editors, *From learning theory to connectionist theory: essays in honor of W.K. Estes*, volume 1, pages 149–168. Erlbaum, Hillsdale, NJ.
- S. Padó and M. Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199.
- M. Palmer, H. Trang Dang, and C. Fellbaum. 2007. Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering*, 13:137–163.
- P. Pantel and D. Lin. 2002. Discovering word senses from text. In *Proceedings of KDD*, Edmonton, Canada.
- R. Passonneau, A. Salieb-Aouissi, V. Bhardwaj, and N. Ide. 2010. Word sense annotation of polysemous words by multiple annotators. In *Proceedings of LREC-7*, Valleta, Malta.
- S. Pradhan, E. Loper, D. Dligach, and M. Palmer. 2007. Semeval-2007 task 17: English lexical sample, SRL and all words. In *Proceedings of SemEvalj*, Prague, Czech Republic.
- L. Pyllkanen, R. Llinas, and G.L. Murphy. 2006. The representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience*, 18:97–109.
- J. Reisinger and R.J. Mooney. 2010. Multi-prototype vector-space models of word meaning. In *Proceeding of NAACL*.
- E. Rosch and C. B. Mervis. 1975. Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605.
- E. Rosch. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104:192–233.
- M. Sahlgren and J. Karlgren. 2005. Automatic bilingual lexicon acquisition using random indexing of parallel corpora. *Journal of Natural Language Engineering, Special Issue on Parallel Texts*, 11(3).
- H. Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1).
- R. Shepard. 1987. Towards a universal law of generalization for psychological science. *Science*, 237(4820):1317–1323.
- E. E. Smith and D. L. Medin. 1981. *Categories and Concepts*. Harvard University Press, Cambridge, MA.
- E. E. Smith, D. Osherson, L. J. Rips, and M. Keane. 1988. Combining prototypes: A selective modification model. *Cognitive Science*, 12(4):485–527.

- P. Smolensky. 1990. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 46:159–216.
- B. Snyder and M. Palmer. 2004. The English all-words task. In *3rd International Workshop on Semantic Evaluations (SensEval-3) at ACL-2004*, Barcelona, Spain.
- J. Taylor. 1989. *Linguistic Categorization: Prototypes in Linguistic Theory*. Oxford Textbooks in Linguistics.
- S. Thater, G. Dinu, and M. Pinkal. 2009. Ranking paraphrases in context. In *Proceedings of the ACL Workshop on Applied Textual Inference*, Singapore.
- D. H. Tuggy. 1993. Ambiguity, polysemy and vagueness. *Cognitive linguistics*, 4(2):273–290.
- G. Vigliocco, D. P. Vinson, W. Lewis, and M. F. Garrett. 2004. Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, 48:422–488.
- W. Weaver. 1949. Translation. In W.N. Locke and A.D. Booth, editors, *Machine Translation of Languages: Fourteen Essays*. MIT Press, Cambridge, MA.