

Supporting inferences in semantic space: representing words as regions

Katrin Erk
University of Texas at Austin
katrin.erk@mail.utexas.edu

Abstract

Semantic space models represent the meaning of a word as a vector in high-dimensional space. They offer a framework in which the meaning representation of a word can be computed from its context, but the question remains how they support inferences. While there has been some work on paraphrase-based inferences in semantic space, it is not clear how semantic space models would support inferences involving hyponymy, like *horse ran* \rightarrow *animal moved*. In this paper, we first discuss what a point in semantic space stands for, contrasting semantic space with Gärdenforsian conceptual space. Building on this, we propose an extension of the semantic space representation from a point to a region. We present a model for learning a region representation for word meaning in semantic space, based on the fact that points at close distance tend to represent similar meanings. We show that this model can be used to predict, with high precision, when a hyponymy-based inference rule is applicable. Moving beyond paraphrase-based and hyponymy-based inference rules, we last discuss in what way semantic space models can support inferences.

1 Introduction

Semantic space models represent the meaning of a word as a vector in a high-dimensional space, where the dimensions stand for contexts in which the word occurs [14, 10, 21, 20]. They have been used successfully in NLP [15], as well as in psychology [10, 13, 16]. Semantic space models, which are induced automatically from corpus data, can be used to characterize the meaning of an *occurrence* of a word in a specific sentence [17, 3] without recourse to dictionary senses. This is interesting especially in the light of the recent debate about the problems of dictionary senses [9, 7]. However,

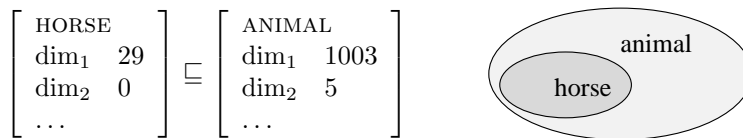


Figure 1: Modeling hyponymy in semantic space: as subsumption between feature structures (left) or as subregion inclusion (right)

it makes sense to characterize the meaning of words through semantic space representations only if these representations allow for inferences.

- (1) Google *acquired* YouTube \implies Google *bought* YouTube
- (2) A *horse ran* \implies An *animal moved*

Ex. (1) is an example of an inference involving a paraphrase: *acquire* can be substituted for *buy* in some contexts, but not all, for example not in contexts involving *acquiring skills*. Ex. (2) is an inference based on *hyponymy*: *run* implies *move* in some contexts, but not all, for example not in the context *computer runs*. In this paper, we concentrate on these two important types of inferences, but return to the broader question of how inferences are supported by semantic space models towards the end.

Semantic space models support paraphrase inferences: Lists of potential paraphrases (for example *buy* and *gain* for *acquire*) and the applicability of a paraphrase rule in context [17] can be read off semantic space representations. The same cannot be said for hyponymy-based inferences. The most obvious conceptualization of hyponymy in semantic space, illustrated in Fig. 1 (left), is to view the vectors as feature structures, and hyponymy as subsumption. However, it seems unlikely that *horse* would occur in a subset of the contexts in which *animal* is found (though see Cimiano et al [1]). There is another possible conceptualization of hyponymy in semantic space, illustrated in Fig. 1 (right): If the representation of a word’s meaning in semantic space were a region rather than a point, hyponymy could be modeled as the sub-region relation. This is also the model that Gärdenfors [5] proposes within his framework of *conceptual spaces*, however it is not clear that the notion of a point in space is the same in conceptual space as in semantic space. To better contrast the two frameworks, we will refer to semantic space as *co-occurrence space* in the rest of this paper.

This paper makes two contributions. First, it discusses the notion of a point in space in both conceptual and co-occurrence space, arguing that they are fundamentally different, with points in co-occurrence space not representing potential entities but mixtures of uses. Second, it introduces a computational model for extending the representation of word meaning

in co-occurrence space from a point to a region. In doing so, it makes use of the property that points in co-occurrence space that are close together represent similar meanings.

We do not assume that the subregion relation will hold between induced hyponym and hypernym representations, no more than that the subsumption relation would hold between them. Instead, we will argue that the region representations make it possible to *encode* hyponymy information collected from another source, for example WordNet [4] or a hyponymy induction scheme [8, 25].

Plan of the paper. Sec. 2 gives a short overview of existing geometric models of meaning. In Sec. 3 we discuss the significance of a point in conceptual space and in co-occurrence space, finding that the two frameworks differ fundamentally in this respect, but that we can still represent word meanings as regions in co-occurrence space. Building on this, Sec. 4 introduces a region model of word meaning in co-occurrence space that can be learned automatically from corpus data. Sec. 5 reports on experiments testing the model on the task of predicting hyponymy relations between occurrences of words. Sec. 6 looks at both paraphrase-based and hyponymy-based inferences to see how their applicability can be tested in co-occurrence space and how this generalizes to other types of inference rules. Sec. 7 concludes.

2 Related work

In this section we give a short overview of three types of geometric models of meaning: co-occurrence space models, conceptual space models, and models of human concept representation.

Co-occurrence space models. Co-occurrence space models (vector space models) represent the meaning of a word as a vector in high-dimensional space [14, 10, 21, 20]. In the simplest case, the vector for a *target word* w is constructed by counting how often each other word co-occurs with w in a given corpus, in a context window of n words around the occurrences of w . Each potentially co-occurring word d then becomes a dimension, and the co-occurrence counts of w with d become the value of w 's vector on dimension d . As an example, Table 1 shows some co-occurrence counts for the target words *letter* and *surprise* in Austen's *Pride and Prejudice*. There are many variations on co-occurrence space representations, for example using syntactic context rather than word co-occurrence [20]. The most important property of co-occurrence space models is that similarity between target words can be estimated as distance in space, using measures such as

	admirer	all	allow	almost	am	and	angry	...
letter	1	8	1	2	2	56	1	...
surprise	0	7	0	0	4	22	0	...

Table 1: Some co-occurrence counts for *letter*, *surprise* in Austen’s *Pride and Prejudice*

Euclidean distance or cosine similarity. Co-occurrence space models have been used both in NLP [15, 25, 22] and in psychology [10, 13, 16]. They have mostly been used to represent the meaning of a word by summing over all its occurrences. We will call these vectors *summation vectors*. A few studies have developed models that represent the meaning of an occurrence of a word in a specific sentence [10, 22, 17, 3]. The *occurrence vector* for a word in a specific sentence is typically computed by combining its summation vector with that of a single context word in the same sentence, for example the direct object of a target verb. For Ex. (1) this would mean computing the meaning representation of this occurrence of *acquire* by combining the summation vectors of *acquire* and *YouTube*. The simplest mechanism that has been used for the vector combination is vector addition. Models for occurrence meaning have typically been tested on a task of judging paraphrase appropriateness: A model is given an occurrence, for example *acquire* in Ex. (1), and a potential paraphrase, for example *buy*. The model then estimates the appropriateness of the paraphrase in the current context as the similarity of the occurrence vector of *acquire* with the summation vector of *buy*.

Conceptual space. Gärdenfors [5] proposes representing concepts as regions in a *conceptual space*, whose *quality dimensions* correspond to interpretable features. In the simplest case, those are types of sensory perception (color, temperature). Gärdenfors defines *natural properties* to be properties that occupy *convex* regions in conceptual space, proposing that all properties used in human cognition are natural. Natural properties offer a solution to the problem of induction if “undesirable” properties such as *grue* [6] do not form convex regions.

Human concept representation. In psychology, feature vector based models of human concept representation (e.g. [24, 19]) are used to model categorization. Since many experiments on human concept representation have been performed using verbal cues, these models represent aspects of word meaning [18], possibly along with other types of knowledge. Nosofsky’s Generalized Context Model (GCM) [19] models the probability of catego-

rizing an exemplar \vec{e} as a member of a concept C as

$$P(C|\vec{e}) = \frac{\sum_{\vec{\eta} \in C} w_{\vec{\eta}} \text{sim}(\vec{\eta}, \vec{e})}{\sum_{\text{concept } C'} \sum_{\vec{\eta} \in C'} w_{\vec{\eta}} \text{sim}(\vec{\eta}, \vec{e})}$$

where the concept C is a set of remembered exemplars, $w_{\vec{\eta}}$ is an exemplar weight, and the similarity $\text{sim}(\vec{\eta}, \vec{e})$ between $\vec{\eta}$ and \vec{e} is defined as $\text{sim}(\vec{\eta}, \vec{e}) = \exp(z \cdot \sum_{\text{dimension } i} w_i (\eta_i - e_i)^2)$. z is a general sensitivity parameter, w_i is a weight for dimension i , and η_i, e_i are the values of $\vec{\eta}$ and \vec{e} on dimension i .

3 Points in co-occurrence space

Since we need to be clear about the entities about which we perform inferences, it is important to understand what a *point* in conceptual and co-occurrence space stands for. This is the topic of this section.

In *conceptual* space, a point is a potential entity, quality, or event. In the region occupied by the concept *yellow*, each point denotes a hue of yellow. In *co-occurrence* space, on the other hand, the representation of *yellow* is a point, the summation vector. corpus occurrences of *yellow*, *yellow door* as well as *yellow pages*. The summation vector is thus not a potential percept, but a sum or mixture of uses. As vectors are computed entirely from observed contexts, summation vectors can be computed for words like *yellow* just as well as *tomorrow* or *idea*. Furthermore, the summation vector is a *representation* of the word’s meaning, rather than a meaning.

An occurrence vector is also a point in co-occurrence space. It, too, does not represent a potential entity. It is computed from two summation vectors: Summation vectors are primary, and occurrence vectors are derived, in all current co-occurrence space approaches. Computing the meaning representation of *acquire* in the context of *YouTube* by combining $\vec{acquire}$ and $\vec{YouTube}$ amounts to constructing a pseudo-summation vector for a word *acquire-in-the-context-of-YouTube*, making pseudo-counts for the dimensions based on the context counts of *acquire* and *YouTube*. If the occurrence vector is computed through addition, as $\vec{acquire} + \vec{YouTube}$, we are basically taking the contexts in which *acquire-in-the-context-of-YouTube* has been observed to be the union of the contexts of *acquire* and *YouTube*. So both summation and occurrence vectors are, in fact, summation vectors representing mixtures of uses. They do not describe potential entities, like points in conceptual space, but are representations of potential meanings of words.

The regions in co-occurrence space we want to identify will thus not be regions of similar entities, but regions of similar mixtures of uses. *Encoding*

external hyponymy information in a co-occurrence space, as we will do below, thus means stating that any mixture of uses in which the hyponym can occur is also a mixture of uses where the hypernym could be found. This is plausible for pairs like *horse* and *animal*, though it stretches corpus reality somewhat for other hypernyms of *horse* like *vertebrate*.

4 A model for regions in co-occurrence space

In this section, we develop a model for automatically inducing region representations for word meaning in co-occurrence space. Our aim is to induce a region representation from existing summation vectors and occurrence vectors. There is existing work on inducing regions from points in a geometric model, in psychological models of human concept representation (Sec. 2). These models use either a single point (prototype models)¹ or a set of points (exemplar models), and induce regions of points that are sufficiently close to the prototype or exemplars. They share two central properties, both of which can be observed in the GCM similarity formula (Sec. 2): **(P1)** Dimensions differ in how strongly they influence classification. **(P2)** Similarity decreases exponentially with distance (Shepard’s law, [23]). We adopt (P1) and (P2) for our model in co-occurrence space. As co-occurrence space can model conceptual phenomena like lexical priming [13], it is reasonable to assume that its notion of similarity matches that of conceptual models. We construct a prototype-style model, with the summation vector as the prototype, using the following additional assumptions: **(P3)** The representation of a word’s meaning in co-occurrence space is a contiguous region surrounding the word’s summation vector. **(P4)** The region includes the occurrence vectors of the word. Property (P4) builds on the argument from Sec. 3 that occurrence vectors are pseudo-summation vectors. It also matches previous work on judging paraphrase appropriateness (Sec. 2), since those studies successfully rely on the assumption that occurrence vectors will be close to summation vectors that represent similar meanings.

We define a model for region representations of word meaning that is based on distance from the summation vector, and that uses the occurrence vectors to determine the distance from the summation vector at which points should still be considered as part of the region. For a given target word w , we construct a log-linear model that estimates the probability $P(\text{IN}|\vec{x})$ that a point \vec{x} is inside the meaning region of w , as follows:

¹Some prototype models use a prototype that has a weighted list of possible values for each feature.

$$P(\text{IN}|\vec{x}) = \frac{1}{Z} \exp\left(\sum_i \beta_i^{\text{IN}} f_i(\vec{x})\right) \quad (1)$$

where the f_i are features that characterize the point \vec{x} , and the β_i^{IN} are weights identifying the importance of the different features for the class IN. Z is a normalizing factor ensuring that the result is a probability. Let $\vec{w} = \langle w_1, \dots, w_n \rangle$ be the summation vector for the word w , in a space of n dimensions. Then we define the features f_i , for $1 \leq i \leq n$, to encode distance from \vec{w} in each dimension, with

$$f_i(\vec{x}) = (w_i - x_i)^2 \quad (2)$$

This log-linear model has property (P1) through the weights β_i^{IN} . It has property (P2) through the exponential relation between the estimated probability and the distances f_i . We will use occurrence vectors of w (as positive data) and occurrence vectors of other words (as negative data) to estimate the β_i during training, thus calibrating the weights by the distances between the summation vector and known members of the region. In the current paper, we will consider only regions with sharp boundaries, which we obtain by placing a threshold of 0.5 on the probability $P(\text{IN}|\vec{x})$. However, we consider it important that this model can also be used to represent regions with soft boundaries by using $P(\text{IN}|\vec{x})$ without a threshold. It may thus be able to model borderline uses of a word, and unclear boundaries between senses [2].

5 Experiments on hyponymy

In this section, we report on experiments on hyponymy in co-occurrence space. We test whether different co-occurrence space models can predict, given meaning representations (summation vectors, occurrence vectors, or regions) of two words, whether one of the two words is a hypernym of the other. In all tests, the models do not see the words, just the co-occurrence space representations.

Experimental setting. We used a Minipar [11] dependency parse of the British National Corpus (BNC) as the source of data for all experiments below. The written portion of the BNC was split at random into two halves: a training half and a test half. We used WordNet 3.0 as the “ground truth” against which to evaluate models. We work with two main sets of lemmas: first, the set of monosemous verbs according to WordNet (we refer to this set as *Mon*), and second, the set of hypernyms of the verbs in *Mon* (we call this set *Hyp*). We concentrate on monosemous words in the current

paper since they will allow us to evaluate property (P3) most directly. Since the model from Sec. 4 needs substantive amounts of occurrence vectors for training, we restricted both sets *Mon* and *Hyp* to verbs that occur with at least 50 different direct objects in the training half of the BNC. The direct objects, in turn, were restricted to those that occurred no more than 6,500 and no less than 270 times with verbs in the BNC, to remove both uninformative and sparse objects. (The boundaries were determined heuristically by inspection of the direct objects for this pilot study.) This resulted in a set *Mon* consisting of 120 verbs, and *Hyp* consisting of 430 verbs. Summation vectors for all words were computed with the `dv` package² from the training half of the BNC, using vectors of 500 dimensions with raw co-occurrence counts as dimension values.

Experiment 1: Subsumption. Above, we have hypothesized that co-occurrence space representations of hyponyms and hypernyms, in the form in which they are induced from corpus data, cannot in general be assumed to be in either a subsumption or a subregion relation. We test this hypothesis, starting with subsumption. We define subsumption as $\vec{x} \sqsubseteq \vec{y} \iff \forall i (y_i > 0 \rightarrow x_i > 0)$. Now, any given verb in *Mon* will be the hyponym of some verbs in *Hyp* and unrelated to others. So we test, for each summation vector \vec{v}_1 of a verb in *Mon* and summation vector \vec{v}_2 of a verb in *Hyp*, whether $\vec{v}_1 \sqsubseteq \vec{v}_2$. The result is that *Mon* verbs **subsume 5% of the *Hyp* verbs of which they are hyponyms**, and **1% of the *Hyp* verbs that are unrelated**. We conclude that subsumption between summation vectors in co-occurrence space is not a reliable indicator of the hyponymy relation between words.

Experiment 2: Subregion relation. Next, we test whether, when we represent a *Hyp*-verb as a region in co-occurrence space, occurrences of its *Mon*-hyponyms fall inside that region, and occurrences of non-hypernyms are outside. First, we compute occurrence vectors for each *Hyp* or *Mon* verb v as described in Sec. 2: Given an occurrence of a verb v , we compute its occurrence vector by combining the summation vector of v with the summation vector of the direct object of v in the given sentence³. We combine two summation vectors by computing their average. In this experiment, we use occurrences from both halves of the BNC. With those summation and occurrence vectors in hand, we then learn a region representation for each *Hyp* verb using the model from Sec. 4. We implemented the region model using the OpenNLP maxent package⁴. Last, we test, for each *Mon* verb

²<http://www.nlpado.de/~sebastian/dv.html>

³Occurrences without a direct object were not used in the experiments.

⁴<http://maxent.sourceforge.net/>

occurrence vector and each *Hyp* region, whether the occurrence vector is classified as being inside the region. The result is that the region models classified zero hyponym occurrences as being inside, resulting in **precision and recall of 0.0**. These results show clearly that our earlier hypothesis was correct: The co-occurrence representations that we have induced from corpus data do not lend themselves to reading off hyponymy relations through either subsumption or the subregion relation.

Experiment 3: Encoding hyponymy. These findings do not mean that it is impossible to test the applicability of hyponymy-based inferences in co-occurrence space. If we cannot *induce* hyponymy relations from existing vector representations, we may still be able to *encode* hyponymy information from a separate source such as WordNet. Note that this would be difficult in a words-as-points representation: The only possibility there would be to modify summation vectors. With a words-as-regions representation, we can keep the summation vectors constant and modify the regions. Our aim in this experiment is to produce a region representation for a *Hyp* verb v such that occurrence vectors of v 's hyponyms will fall into the region. We use only *direct* hypernyms of *Mon* verbs in this experiment, a 273-verb subset of *Hyp* we call **DHyp**. For each *DHyp* verb v , we learn a region representation centered on v 's summation vector, using as positive training data all occurrences of v and v 's *direct hyponyms* in the training half of the BNC. (Negative training data are occurrences of other *DHyp* verbs and their children.) We then test, for each occurrence of a *Mon* verb in the test half of the BNC that does not occur in the training half with the same direct object, whether it is classified as being inside v 's region. The result of this experiment is a **precision of 95.2, recall of 43.4, and F-score of 59.6** (against a random baseline of prec=11.0, rec=50.2, and F=18.0). This shows that it is possible to encode hyponymy information in a co-occurrence space representation: The region model identifies hyponym occurrences with very high precision. If anything, the region is too narrow, classifying many actual hyponyms as negatives.

6 Inference in co-occurrence space

In this section we take a step back to ask what it means for co-occurrence space to support inferences, taking the inferences in Ex. (1) and (2) as an example. The inference in Ex. (1), which involves a paraphrase, is supported in two ways: **(I1)** Paraphrase candidates – words that may be substituted for *acquire* in some contexts – can be read off a co-occurrence space represen-

tation [12]. They are the words whose summation vectors are closest to the summation vector of *acquire* in space. In this way, co-occurrence space can be used for the *construction* of context-dependent paraphrase rules. **(I2)** Given an occurrence \vec{o} of *acquire*, the appropriateness of applying the paraphrase rule substituting *buy* for *acquire* is estimated based on the distance between \vec{o} and the summation vector \vec{b} of *buy* [17, 3]. This can be used to select the single best paraphrase candidate for the given occurrence of *acquire*, or to produce a ranking of all paraphrase candidates [3].

Concerning the hyponymy-based inference in Ex. (2), we have established in Experiments 1 and 2 (Sec. 5) that it is at least not straightforward to construct hyponymy-based rules from co-occurrence space representations in analogy to (I1). However, **(H2)** Experiment 3 has shown that, given a set of attested occurrences of hyponyms of *move*, we can construct a region representation for *move* in co-occurrence space that can be used to test applicability of hyponymy-based rules: The appropriateness of applying the hyponymy-based rule substituting *move* for this specific occurrence of *run* can be estimated based on whether the occurrence vector of *run* is located inside the *move* region.

In both (I2) and (H2), an inference rule is “attached” to a point or a region in co-occurrence space: the summation vector of *buy* in (I2), the region representation of *move* in (H2). The inference rule is considered applicable to an occurrence if its occurrence vector is close enough in space to the *attachment point* or inside the *attachment region*. Co-occurrence space thus offers a natural way of determining the applicability of a (paraphrase or hyponymy-based) inference rule to a particular occurrence, via distance to the attachment point or inclusion in the attachment region. Applicability can be treated as a yes/no decision, or it can be expressed through a graded degree of confidence. In the case of attachment *point*, this degree of confidence would simply be the similarity between occurrence vector and attachment point. Concerning attachment *regions*, note that the model of Sec. 4 actually estimates a probability of region inclusion for a given point in space. In this paper, we have placed a threshold of 0.5 on the probability to derive hard judgments, but the probability can also be used directly as a degree of confidence.

The general principle of using co-occurrence space representation to rate inference rule applicability, and to do this by linking rules to attachment points or regions, could maybe be used for other kinds of inference rules as well. The prerequisite is, of course, that it must make sense to judge rule applicability through a single attachment point or region.

7 Conclusion and outlook

In this paper, we have studied how semantic space representations support inferences, focusing on hyponymy. To encode hyponymy through the sub-region relation, we have considered word meaning representations through regions in semantic space. We have argued that a point in semantic space represents a mixture of uses, not a potential entity, and that the regions in semantic space we want to identify are those that represent the same or similar meanings. We have introduced a computational model that learns region representations, and we have shown that this model can predict hyponymy with high precision. Finally, we have suggested that semantic space supports inferences by attaching inference rules to points or regions in space and licensing rule application depending on distance in space. It is an open question how far the idea of attachment points and attachment regions can be extended beyond the paraphrase and hyponymy rules we have considered here; this is the question we will consider next.

Acknowledgements. Many thanks to Manfred Pinkal, Jason Baldridge, David Beaver, Graham Katz, Alexander Koller, and Ray Mooney for very helpful discussions. (All errors are, of course, my own.)

References

- [1] P. Cimiano, A. Hotho, and S. Staab. Learning concept hierarchies from text corpora using formal concept analysis. *Journal of Artificial Intelligence Research*, 24:305–339, 2005.
- [2] K. Erk and S. Padó. Towards a computational model of gradience in word sense. In *Proceedings of IWCS-7*, Tilburg, The Netherlands, 2007.
- [3] K. Erk and S. Pado. A structured vector space model for word meaning in context. In *Proceedings of EMNLP-08*, Hawaii, 2008.
- [4] C. Fellbaum, editor. *WordNet: An electronic lexical database*. MIT Press, Cambridge, MA, 1998.
- [5] P. Gärdenfors. *Conceptual spaces*. MIT press, Cambridge, MA, 2004.
- [6] N. Goodman. *Fact, Fiction, and Forecast*. Harvard University Press, Cambridge, MA, 1955.
- [7] P. Hanks. Do word meanings exist? *Computers and the Humanities*, 34(1-2):205–215(11), 2000.
- [8] M. Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of COLING 1992*, Nantes, France, 1992.

- [9] A. Kilgarriff. I don't believe in word senses. *Computers and the Humanities*, 31(2):91–113, 1997.
- [10] T. Landauer and S. Dumais. A solution to Platos problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240, 1997.
- [11] D. Lin. Principle-based parsing without overgeneration. In *Proceedings of ACL'93*, Columbus, Ohio, USA, 1993.
- [12] D. Lin. Automatic retrieval and clustering of similar words. In *COLING-ACL98*, Montreal, Canada, 1998.
- [13] W. Lowe and S. McDonald. The direct route: Mediated priming in semantic space. In *Proceedings of the Cognitive Science Society*, 2000.
- [14] K. Lund and C. Burgess. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, 28:203–208, 1996.
- [15] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [16] S. McDonald and M. Ramscar. Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. In *Proceedings of the Cognitive Science Society*, 2001.
- [17] J. Mitchell and M. Lapata. Vector-based models of semantic composition. In *Proceedings of ACL-08*, Columbus, OH, 2008.
- [18] G. L. Murphy. *The Big Book of Concepts*. MIT Press, 2002.
- [19] R. M. Nosofsky. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115:39–57, 1986.
- [20] S. Padó and M. Lapata. Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199, 2007.
- [21] M. Sahlgren and J. Karlgren. Automatic bilingual lexicon acquisition using random indexing of parallel corpora. *Journal of Natural Language Engineering, Special Issue on Parallel Texts*, 11(3), 2005.
- [22] H. Schütze. Automatic word sense discrimination. *Computational Linguistics*, 24(1), 1998.
- [23] R. Shepard. Towards a universal law of generalization for psychological science. *Science*, 237(4820):1317–1323, 1987.
- [24] E. E. Smith, D. Osherson, L. J. Rips, and M. Keane. Combining prototypes: A selective modification model. *Cognitive Science*, 12(4):485–527, 1988.
- [25] R. Snow, D. Jurafsky, and A. Y. Ng. Semantic taxonomy induction from heterogenous evidence. In *Proceedings of COLING/ACL'06*, 2006.