Grounded Distributional Semantics for Abstract Words

Katsumi Takano (katsumi@utm.inf.uec.ac.jp)

Akira Utsumi (utsumi@uec.ac.jp)

Department of Informatics, The University of Electro-Communications
1-5-1, Chofugaoka, Chofushi, Tokyo 182-8585, Japan

Abstract

Since Harnad (1990) pointed out the symbol grounding problem, cognitive science research has demonstrated that grounding in perceptual or sensorimotor experience is crucial to language. Recent embodied cognition theories have argued that language is more important for grounding abstract than concrete words; abstract words are grounded via language. Distributional semantics has recently addressed the embodied nature of language and proposed multimodal semantic models. However, these models are not cognitively plausible because they do not address the recent embodiment view of abstract concepts. Therefore, we propose a novel multimodal distributional semantics in which abstract words are represented indirectly through grounded representations of their semantically related concrete words. A simulation experiment demonstrated that the proposed model achieved better performance in computing the word similarity than other multimodal or text-based distributional models. This finding suggests that the indirect embodiment view is plausible and contributes to the improvement of multimodal distributional semantics.

Introduction

Distributional semantics is a computational approach to constructing semantic representations of words by observing distributional statistics of word occurrence in large collections of text. The lexical meaning of a word is represented by a high-dimensional vector in a semantic space, and the degree of semantic relatedness between any two words can be easily computed from their vectors, for example as the cosine of two vectors. Because of its simplicity and versatility, distributional semantics has recently received enormous attention, and a variety of methods have been proposed and refined (Turney & Pantel, 2010; Jones, Willits, & Dennis, 2015). Recent advances of neural network language models or word embedding models (Mikolov, Chen, Corrado, & Dean, 2013) are also regarded as notable approaches of distributional semantics.

However, distributional semantics has been criticized as psychologically implausible because it is ungrounded or disembodied (Glenberg & Robertson, 2000). Embodied cognition theory argues that language is grounded in our sensory perception and experience, but distributional semantic models learn from only linguistic information and do not use sensorimotor information. Although it is controversial whether distributional semantics cannot essentially simulate human semantic memory, it is no doubt that it cannot represent the meaning of some kinds of words, in particular concrete words, just as they are represented in human semantic memory.

For this reason, recent studies of distributional semantics have addressed the use of sensorimotor or perceptual information to learn psychologically plausible lexical representation. For example, Andrews, Vigliocco, and Vinson (2009) used perceptual features collected as featural norm to ground a language-based topic model on perceptual experience, and demonstrated that the integrated model outperformed the language-based topic model. Although the featural norm used in their study is represented by language, recent attempts have been directed toward multimodal distributional semantics (for a review, see Baroni, 2016), in which textual information is integrated with multimodal information computed directly from nonlinguistic inputs such as visual (Bruni, Tran, & Baroni, 2014; Kiela, Hill, Korhonen, & Clark, 2014), auditory (Kiela & Clark, 2015), or olfactory (Kiela, Bulat, & Clark, 2015) ones.

Their primary research purpose is to improve the accuracy of distributional semantic models by using multimodal information and it has often been demonstrated that multimodal information can improve the performance of language-based distributional semantics. However, this is not always the case. Kiela et al. (2014) showed that, although the inclusion of visual input improves the representations of concrete words, it degrades the representations of abstract words. This finding is highly important, because it seems to suggest that abstract words may not be embodied in the same way as concrete words are embodied. It has also been pointed out by recent development of the embodied view of language (e.g., Louwerse, 2011; Borghi & Cangelosi, 2014; Thill, Padó, & Ziemke, 2014). These views argue that abstract words are embodied differently from concrete words, and acquisition of abstract words depends more heavily on linguistic experience. To provide a psychologically more plausible model, distributional semantics must deal with the problem of abstract words, and at the same time, such a theory can provide an effective computational framework for testing the validity of the embodiment theory of language.

In this paper, therefore, we propose a novel distributional semantic model by taking into account the recent embodied view of abstract concepts. We focus on visual images as a source of perceptual information, as used in many other studies on multimodal distributional semantics, and propose a mechanism of embodiment of abstract concepts via language, in which an embodied vector representation of an abstract word is obtained from the embodied representations of concrete words that are semantically related to that abstract word. We then test the psychological plausibility of the proposed method using human similarity ratings of English word pairs.

Embodiment of Abstract Words

Abstract concepts pose a serious challenge to the embodied theory of language. As Dove (2015) pointed out, it is difficult to see how representations grounded in perceptual experience can capture the content of abstract concepts or words such as

truth and *justice*. Even if abstract concepts somewhat depend on perceptual experience, the way in which such abstract concepts are grounded seems to differ from concrete concepts. This intuition has been supported by a number of empirical findings obtained in cognitive neuroscience. Neuroimaging studies have demonstrated that processing of abstract concepts elicits greater activation of the left perisylvian language network (such as the left inferior frontal gyrus and the left superior temporal cortex) as compared to processing of concrete concepts (e.g., Binder, Desai, Graves, & Conant, 2009; Wang, Conder, Blitzer, & Shinkareva, 2010). Vigliocco et al. (2014) also found greater engagement of the rostral anterior cingulate cortex, an area associated with emotional processing, for abstract than concepts words.

Some embodied theories claim a general mechanism of embodiment common to both concrete and abstract concepts. For example, Barsalou's (1999) theory of perceptual symbol systems argues that concepts (and meanings as well) are not amodal, formal symbols, but rather inherently modal, perceptual symbols, which refer to perceptual states (or neural representations) in sensorimotor systems. These perceptual symbols become integrated into a simulator for a concept that enables us to simulate sensorimotor experience of perceiving and/or acting on exemplars of that concept. According to Barsalou (1999), abstract concepts are represented in the same perceptual symbol systems, although the simulator of abstract concepts is constructed from not only sensorimotor but also introspective states or events. Barsalou (1999) explains this by using examples such as *truth* and *negation*, but the explanation is relatively vague and limited to some logical or affective concepts.

Recently, a number of studies have attempted to resolve the problem of abstract concepts while maintaining the embodied view by emphasizing the role of language in processing of abstract concepts. For example, Word-as-Tools (WAT) theory (Borghi & Cangelosi, 2014) claims that words function as tools to extend our cognition and such the linguistic function is more crucial for abstract than for concrete words representation. Dove (2014) argues for a similar view that language provides an important means of extending our cognitive or mental capabilities by enabling access to an embodied representational system that exists independently of language. The basic idea underlying these views is that abstract words are grounded in sensorimotor or perceptual experiences, but the embodiment is indirect, rather than direct in the case of concrete words. This "indirect embodiment" view is proposed more clearly by Thill et al. (2014). They propose a "division of labor" approach between a perceptual layer that associates basic, concrete concepts with perceptual features and a relational layer that grounds more complex and abstract concepts in relation to basic concepts. Gleitman et al.'s (2005) explanation of how "hard words" are acquired is closely related to the indirect embodiment view. In the early stage of lexical acquisition, the meaning of concrete words is acquired directly from perceptual information via word-to-world pairing. In the later stage, the meaning of hard words, which is not easily accessible through perception, is acquired by a structure-

to-world mapping procedure that combines linguistic observations with cooccurring perceptual experience. From the distributional semantics side, Louwerse (2011) also put forward the indirect embodiment view, namely the symbol interdependency hypothesis, to refute a criticism against distributional semantics made by the embodied view. According to the symbol interdependency hypothesis, language comprehension is symbolic through interdependencies of amodal linguistic symbols, while it is indirectly embodied through the references linguistic symbols make to perceptual representations. Hence, "language has evolved to become a communicative short-cut for language users and encodes relations in the world, including embodied relations (*ibid.*, p.279)."

However, despite receiving much attention from cognitive science over the recent years, the indirect embodiment view of abstract words has not been addressed by computational studies on distributional semantics, as pointed out in the introduction. Although multimodal distributional semantics incorporates the embodied view of language into an algorithm for constructing word vectors, it cannot deal with the problem of abstract words. Kiela et al.'s (2014) approach clearly shows this difficulty; to improve the accuracy of multimodal word vectors, they excluded visual information when constructing vectors of abstract words, while including visual information for concrete words. For distributional semantics to be a psychologically more plausible model, it is highly necessary to integrate the indirect embodiment view into multimodal distributional semantics. This is what we aim to accomplish in this paper.

Grounded Distributional Semantics

Our grounded distributional semantic model, whose algorithm is shown in Figure 1, is based on the indirect embodiment view that grounding of abstract words is mediated by language. Therefore, different grounding methods are used for concrete and abstract words. For a concrete word, its vector representation is constructed by combining a text-based vector (computed from a text corpus) with a visual vector computed from images for that word (i.e., Steps 3(a) and 4 of Figure 1). This method does not differ from Kiela et al.'s (2014) and other standard methods for multimodal distributional semantics. On the other hand, the vector representation of an abstract word is constructed by combining a text-based vector with visual vectors of concrete words semantically related to that abstract word (i.e., Steps 3(b) and 4).

An important part of this algorithm is the classification between concrete and abstract words at Step 2. In this paper, we use the image dispersion method proposed by Kiela et al. (2014). This method is motivated by the intuition that the diversity of images retrieved for a particular word depends on its concreteness. It is generally expected that images for concrete concepts such as "*apple*" are more similar to one another than those for abstract concepts such as "*happiness*." The degree of image dispersion for a concept is defined as the average pairwise cosine similarity between all the vector representations in the set of images for that concept. All words are sorted in descending order of image dispersion, and top

1. Construct a linguistic semantic space DSM_L from a text corpus.

2. Divide the vocabulary V into concrete words V_C and abstract words V_A (i.e., $V = V_C \cup V_A$ and $V_C \cap V_A = \phi$).

3. Construct a visual semantic space DSM_V using visual images.

   (a) For a concrete word $w_i^C \in V_C$, compute a visual vector $\boldsymbol{y}(w_i^C)$ directly from the images labeled with w_i^C.

   (b) For an abstract word $w_i^A \in V_A$, compute a visual vector $\boldsymbol{y}(w_i^A)$ as $\sum_{w_j^C \in SN(w_i^A)} \boldsymbol{y}(w_j^C)$ where $SN(w_i^A) \subset V_C$ is a set of n semantic neighbors of (i.e., n concrete words semantically related to) w_i^A. Semantic neighbors of w_i^A are computed in the linguistic semantic space DSM_L.

4. Construct a grounded semantic space DSM_G by combining DSM_L and DSM_V as follows. For each word $w_i \in V$, normalize $\boldsymbol{x}(w_i) \in DSM_L$ and $\boldsymbol{y}(w_i) \in DSM_V$, and concatenate them.

Figure 1: Overall algorithm of the proposed distributional semantic model.

$100 p_{abs}$ $(0 < p_{abs} < 1)$ percent are judged to be abstract and constitute V_A. Kiela et al. (2014) confirmed that the degree of image dispersion is a valid measure for concreteness by showing a high correlation between concreteness ratings and dispersion values. Furthermore, as compared to concreteness rating data, image dispersion is more appropriate for judging the importance of embodiment in the acquisition of word meanings, because the consistency among images for a word is likely to imply that this word refers to a discernible object or action in the world that we can easily recognize through perceptual experience.

Determining semantic neighbors (or in other words, "mediator words") of an abstract word is also a key step for successful modeling. We apply a simple method that chooses n concrete words most similar to a target abstract word. However, this simple method may not work well in some cases. Some highly abstract words (e.g., *truth*, *wisdom*) may not have semantically similar concrete words, but this method forces such abstract words to be represented by visual vectors of n concrete words, which may be relatively more similar than other concrete words but absolutely dissimilar to the abstract words. Hence, we consider another method, in which $N(> n)$ semantic neighbors of an abstract word are selected from the whole vocabulary V, and then n most concrete words are selected from N semantic neighbors according to the degree of image dispersion. If n concrete words cannot be selected (i.e., N semantic neighbors contain less than n concrete words), no visual vector is attached to that abstract word in the same way as Kiela et al.'s (2014) method.

Evaluation Experiment

Test Data

As test data for evaluating distributional semantic models, we used two word similarity norms, namely WordSim-353 (Finkelstein et al., 2001) and SimLex-999 (Hill, Reichart, & Korhonen, 2015), which are well-known gold-standards widely used in the computational linguistics community. These data contain English word pairs (353 pairs in WordSim-353 and 999 pairs in SimLex-999) with human similarity ratings. The similarity ratings in WordSim-353 are regarded as measures of semantic relatedness (or in other words, associative similarity), while the ratings in SimLex-999 reflect semantic similarity (or taxonomic similarity). For example, a pair of *clothes – closet* is rated 8.00 in WordSim-353, but 1.96 in SimLex-999. These two words are related because the word "closet" refers to a small room or cabinet for hanging "clothes," but their meanings are not similar.

Multimodal Semantic Space

We generated a text-based semantic space (i.e., DSM_L in Figure 1) from the British National Corpus with 100M word tokens as follows. First, we constructed a word-cooccurrence matrix with 43,528 different words (and thus the dimensions of the matrix were $43,528 \times 43,528$). The cooccurrence frequency was counted within a sentence. Second, the word-cooccurrence matrix was weighted by positive pointwise mutual information (PPMI), which generally achieves better performance (Bullinaria & Levy, 2007). Finally, the weighted matrix was smoothed by singular value decomposition (SVD) and the dimension of row word vectors was reduced to 300.

To construct a visual semantic space (i.e., DSM_V in Figure 1), we collected 30 images for each word using *Flickr* image retrieval, from which a visual vector for that word was constructed. As a method for visual vector computation, we used the well-known Bag of Visual Words (BoVW) approach (Sivic & Zisserman, 2003). BoVW computes a vector representation for each image by clustering into visual words a set of local descriptors obtained from the image dataset, and then by computing a histogram of visual word counts. In this paper, we extracted SURF descriptors (Bay, Tuytelaars, & Van Gool, 2006) using a dense keypoint sampling method. Each local descriptor is represented by a 3×128-dimensional vector, because a SURF feature is normally represented in a 128-dimensional vector and computed for each of the three components of HSV. All the SURF feature vectors were clustered into 300 visual words, and each image is represented by a 300-dimensional histogram vector of visual words. Finally, a visual vector for each word is computed by summing the visual vectors of all the images retrieved by that word, and then by weighting the summed vector by PPMI.

Method

According to the proposed method shown in Figure 1, we constructed several grounded semantic spaces by changing the values of the parameters p_{abs} (from 0.10 to 0.95 in steps of 0.05), n (from 1 to 5), and N (from 10 to 50 in steps of 10, 100, and all). To evaluate the relative performance of the proposed model, we also generated the standard multimodal representations and Kiela et al.'s (2014) filtering-based representations. In the standard method, every word is represented by concatenating its linguistic vector with a visual vector computed directly from images of that word. In Kiela

Table 1: Correlation coefficients between the cosine similarity computed by the distributional semantic models and the pairwise similarity ratings of two datasets. For the proposed model, the highest correlation is shown boldfaced for each of the two datasets. For the other models (including the proposed model with N =all), results obtained with the same parameters as the proposed model are selectively shown.

Model	WordSim-353	SimLex-999
Text-based DSM_L	0.525	0.248
Visual DSM_V, Standard	0.227	0.118
Visual DSM_V, Proposed ($p_{abs} = 0.50, n = 2, N$ =all)	0.257	0.112
Visual DSM_V, Proposed ($p_{abs} = 0.95, n = 2, N$ =all)	0.185	0.085
Grounded DSM_G, Standard	0.476	0.235
Grounded DSM_G, Kiela et al. (2014) ($p_{abs} = 0.50$)	0.510	0.262
Grounded DSM_G, Kiela et al. (2014) ($p_{abs} = 0.95$)	0.532	0.245
Grounded DSM_G, Proposed ($p_{abs} = 0.50, n = 2, N$ =all)	0.478	0.220
Grounded DSM_G, Proposed ($p_{abs} = 0.95, n = 2, N$ =all)	0.390	0.179
Grounded DSM_G, Proposed ($p_{abs} = 0.95, n = 2, N = 30$)	**0.551**	0.260
Grounded DSM_G, Proposed ($p_{abs} = 0.80, n = 5, N = 30$)	0.535	**0.278**

Note. p_{abs} = the proportion (i.e., threshold) of abstract words; n = the number of semantic neighbors (i.e., mediator words) from which visual vectors of abstract words are computed; N = the size of population from which n mediator words are chosen.

et al.'s (2014) filtering-based method, concrete words, which are determined by image dispersion, are represented in the same way as the standard method (i.e., concatenation of their linguistic and visual vectors), while abstract words are represented only by their linguistic vectors.

The performance of each semantic space was measured by Pearson's correlation coefficient between the computed cosine values and the semantic similarity ratings in the test data. If the proposed semantic space can improve the performance as compared to the standard multimodal and Kiela et al.'s (2014) semantic spaces as well as to the text-based semantic space, it would demonstrate that the indirect embodiment view of abstract words is supported computationally.

Result

Table 1 shows the correlation coefficients between the cosine similarity computed by the distributional semantic models and the similarity ratings of the test data. Overall, the proposed grounded semantic space achieved better performance on both datasets than the text-based model, standard multimodal model, and Kiela et al.'s (2014) model. This result indicates that the proposed model is made psychologically more plausible by language-mediated visual representations of abstract words, thus providing computational evidence for the indirect embodiment view of abstract words.

The result that the standard multimodal model did not outperform the text-based model is consistent with Kiela et al.'s (2014) result; this indicates that inclusion of perceptual input does not always improve the quality of lexical meaning representations. To overcome this difficulty, Kiela et al. (2014) did not use perceptual information to represent image-dispersed words, but we use language-mediated perceptual information for those words, in accordance with the recent idea of embodiment of abstract concepts. Hence, the observed better performance of our model against Kiela et al.'s (2014) model demonstrates the plausibility of our refined representations and the indirect embodiment view.

On the other hand, the proposed method yielded worse performance when all abstract words are represented using the visual vectors of their semantic neighbors (i.e., in the case of N = all in Table 1). This result implies that indirect embodiment via language is not always effective; Kiela et al.'s (2014) method of not including visual information works better in some cases. In other words, some abstract words (e.g., *truth*, *wisdom*) need not to be grounded in perceptual experience and their meanings may be able to be captured primarily by linguistic information.

The proposed grounded model achieved the best performance at very high p_{abs} ($p_{abs} = 0.95$ for WordSim-353 and $p_{abs} = 0.80$ for SimLex-999). This result indicates that only a relatively small number of words need to be grounded directly in perceptual information; most words are likely to be grounded via language or need not to be grounded. It may also implies that word concreteness does not necessarily determine the necessity of grounding in perceptual experience. Table 2 lists some example words that are computed as concrete and abstract by the proposed model. Some highly concrete words (e.g., baby, computer, seafood) whose concreteness ratings (1 to 5) (Brysbaert, Warriner, & Kuperman, 2014) are higher than 4.5 and many moderately concrete words (e.g., currency, interview, summer) whose concreteness ratings are higher than 3.0 are classified as abstract. Nevertheless, indirect grounding is effective in representing these words. The proposed model improved the accuracy of similarity computation for word pairs in WordSim-353 such as *governor–interview*, *currency–market*, *summer–drought*,

Table 2: Examples of concrete and abstract words in WordSim-353 classified by the proposed DSM_G ($p_{abs} = 0.95$, $n = 2$, $N = 30$). For an abstract word, its mediator words (i.e., concrete words used for computing its visual vectors) are also shown in parentheses. Note that no mediator words are shown for some abstract words because N semantic neighbors of those words do not include n concrete words.

Concrete words in V_C

animal, bread, car, carnivore, collection, food, forest, impartiality, jaguar, mammal, marathon, market, moon, office, shower, television, tiger, water, zoo

---

Abstract words in V_A

activity, alcohol, astronomer (*lunar, saturn*), baby, bird, book (*collection, print*), center, chance, children (*playgroup, whist*), cognition (*representational, intentionality*), computer, currency (*ecu, lira*), discovery, drought (*soil, rainforest*), experience, family (*home, bear*), flood (*stream, storm*), fruit (*bean, salad*), girl, governor (*governorship, appoint*), index, interview (*quote, television*), king, law, liquid (*water, bubble*), love, marriage, media (*magazine, television*), museum (*collection, painting*), music, psychology, reason, seafood (*salad, pasta*), sprint (*marathon, athlete*), summer (*garden, lake*), treatment, word

---

summer:

garden:

lake:

Figure 2: Examples of images for an indirectly grounded word *summer* and its mediator words *garden* and *lake* used in our multimodal distributional semantic model.

and *seafood–food*, as compared to the text-based model and the standard multimodal model. For example, as shown in Figure 2, a variety of images associated with the word *summer* seem not to represent one discernible concept, and thus they are somewhat harmful for an appropriate visual representation of *summer*. On the other hand, images for the mediator words *garden* and *lake* are highly consistent, and these visual information can represent some aspects of summer.

Hence, it may be inappropriate that we refer to the partitioning of vocabulary (V_A and V_C at Step 2 of Figure 1) in terms of abstractness. Rather, what this partitioning really reflects is the necessity of grounding in perceptual experience. Note that this finding is consistent with the existing findings on lexical acquisition (e.g., Gleitman et al., 2005). "Easy" words (mostly nouns), which express concrete basic-level concepts, are learned at the early stage of word learning,

and then many other words are learned through the knowledge of easy words, leading to a rapid vocabulary growth.

Discussion

In this paper, inspired by the recent view of the embodiment of abstract words, we have proposed a novel approach to multimodal distributional semantics. In the proposed method, the grounded visual representation of an abstract word is learned through semantically related mediator words, rather than directly from images for that word as used in the standard multimodal distributional semantics. The plausibility of the proposed method is tested and justified by comparing the cosine similarity computed by distributional semantics with human semantic similarity ratings. This finding can be interpreted as supporting the indirect embodiment view of abstract words.

However, our model is not so sophisticated and rather preliminary. It did not achieve a statistically significant improvement over the existing multimodal semantic models. We can point out several issues to improve the quality of the model. One important issue to address for future research is a method for choosing mediator words whose visual vectors form the grounded representation of an abstract word (at Step 3(b) of Figure 1). In this paper, we simply apply semantic similarity computed by the text-based model as a measure of choosing mediator words, but a number of empirical studies have suggested that some other factors affect the embodiment of abstract words. For example, emotion (or affect) is suggested to play an important role in processing abstract words (Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Vigliocco et al., 2014). Affective experience is crucial in the grounding of abstract concepts, and thus it is likely that affectively similar words are effective mediators of embodiment for abstract words. Age-of-acquisition may also be related to the embodiment of abstract words, because basic words learned at the early stage of lexical acquisition are represented primarily perceptually, while other words learned at the later stage are acquired through the knowledge of basic words (Gleitman et al., 2005; Thill et al., 2014).

Another important research topic is the use of motor information for distributional semantics. Multimodal semantic spaces learned from static images are less than grounded in motor experience. To overcome this difficulty, we can consider distributional semantics using feature vectors extracted from videos or continuous image sequences.

Further refinement can be achieved in the proposed model if we consider different ways of indirect grounding according to the degree of abstractness for words. For example, the number of mediator words can be determined depending on the abstractness of words. We can also combine an original visual vector directly computed from the images for an abstract word with those for mediator words if the original vector is somewhat useful for representing the meaning of the abstract word.

Through these improvements, we can obtain a psychologically more plausible distributional model, and multimodal distributional semantics would hold promise as the computational basis for a psychologically testable theory of embodied

cognition and symbol grounding.

References

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*, 463–498.

Baroni, M. (2016). Grounding distributional semantics in the visual world. *Linguistic Issues in Language Technologies*, *10*(1), 3–13.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features. In A. Leonardis, H. Bischof, & A. Pinz (Eds.), *Computer vision — ECCV 2006* (pp. 404–417). Springer.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*, 2767–2796.

Borghi, A. M., & Cangelosi, A. (2014). Action and language integration: From humans to cognitive robots. *Topics in Cognitive Science*, *6*, 344–358.

Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. *Jounral of Artificial Intelligence Research*, *49*, 1–47.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–991.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*(3), 510–526.

Dove, G. (2014). Thinking in words: Language as an embodied medium of thought. *Topics in Cognitive Science*, *6*, 371–389.

Dove, G. (2015). Three symbol ungrounding problems: Abstract concepts and the future of embodied cognition. *Psychonomic Bulletin & Review*, Online First Articles.

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., & Ruppin, E. (2001). Placeing search in context: The concept revisited. In *Proceedings of the 10th international conference on world wide web* (pp. 406–414).

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64.

Glenberg, A., & Robertson, D. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*, 379–401.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, *42*, 335–346.

Hill, F., Reichart, R., & Korhonen, A. (2015). SimLex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, *41*(4), 665–695.

Jones, M. N., Willits, J., & Dennis, S. (2015). Models of semantic memory. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford handbook of mathematical and computational psychology* (pp. 232–254). New York, NY: Oxford University Press.

Kiela, D., Bulat, L., & Clark, S. (2015). Grounding semantics in olfactory perception. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing* (pp. 231–236).

Kiela, D., & Clark, S. (2015). Multi- and cross-modal semantics beyond vision: Grounding in auditory perception. In *Proceedings of the 2015 conference on empirical methods in natural language processing (emnlp2015)* (pp. 2461–2470).

Kiela, D., Hill, F., Korhonen, A., & Clark, S. (2014). Improving multi-modal representations using image dispersion: Why less is sometimes more. In *Proceedings of the 52nd annual meeting of the association for computational linguistics* (pp. 835–841).

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*, 273–302.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of workshop at the international conference on learning representation (iclr)*.

Sivic, J., & Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *Proceedings of ieee international conference on computer vision (ICCV 2003), volume 2* (pp. 1470–1477).

Thill, S., Padó, S., & Ziemke, T. (2014). On the importance of a rich embodiment in the grounding of concepts: Perspectives from embodied cognitive science and computational linguistics. *Topics in Cognitive Science*, *6*, 545–558.

Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188.

Vigliocco, G., Kousta, S.-T., Rosa, P. A. D., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex*, *24*, 1767–1777.

Wang, J., Conder, J. A., Blitzer, D. N., & Shinkareva, S. V. (2010). Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human Brain Mapping*, *31*, 1459–1468.