An Instance Theory of Distributional Semantics

Randall K. Jamieson (randy.jamieson@umanitoba.ca)
Department of Psychology
University of Manitoba, Winnipeg, MB, Canada

Brendan T. Johns (btjohns@buffalo.edu)
Department of Communicative Disorders and Sciences
University at Buffalo SUNY, Buffalo, NY, USA

Johnathan E. Avery (averjo@iu.edu)
Department of Psychological and Brain Sciences
Indiana University, Bloomington, IN, USA

Michael N. Jones (jonesmn@iu.edu)
Department of Psychological and Brain Sciences
Indiana University, Bloomington, IN, USA

Abstract

Abstraction to a single prototypical representation is a core principle of Distributional Semantic Models (DSMs) that learn semantic representations for words by applying dimension reduction to statistical redundancies in language. While the learning mechanisms for semantic abstraction vary widely across the many DSMs in the literature, they are essentially all prototype models in that they create a single abstract representation for a word's meaning. The prototype method stands in stark contrast to work in the field of categorization that has converged on the importance of instance models. In comparison to the prototype method, instance-based models assume only an episodic store and, rather than applying abstraction mechanisms at learning, argue that meaning emerges in the act of retrieval. We cash this idea out by presenting and evaluating an instance theory of distributional semantics, and by demonstrating that it can explain diverging patterns of homonymous words that classic "abstraction-at-learning" models simply cannot as a consequence of their architectural assumptions.

Keywords: Semantic memory; Instance theory; Latent Semantic Analysis

Introduction

Distributional semantic models (DSMs) such as BEAGLE, HAL, LSA, and Word2Vec represent a major advance in the field of semantic memory (Jones & Mewhort, 2007; Lund & Burgess, 1996; Landauer & Dumais, 1997; Mikolov, et al., 2013). DSMs attempt to explain how humans transform first-order statistical experience with language into deep knowledge representations of word meaning. The mechanisms they posit for this transformation vary widely, ranging from simple co-occurrence counting to reinforcement learning (see Jones, Willits, & Dennis, 2015 for a review). But virtually all DSMs share one commonality: They are prototype models. This shared feature may represent a significant architectural flaw in DSMs, leading the field to assume that abstraction is a learning rather than retrieval mechanism.

All current spatial DSMs use the co-occurrence regularities of words across contexts in language, and attempt to build a single vector representation that best represents the word's aggregate meaning, formalizing the classic notion that "you shall know a word by the company it keeps" (Firth, 1957). However, the notion of building a single prototypical center of tendency is in stark contrast to the current state-of-the-art in related fields, such as categorization and episodic memory. The categorization literature, for example, has long recognized the importance of instance-based models for understanding category knowledge and the contextual disambiguation of meaning; especially given that prototype theories of categorization are simply unable to explain human behavior when dealing with category structures that have non-linearly separated structure, such as in classic XOR.

Jones (2018) has recently suggested that current abstraction-at-learning DSMs suffer from the same issues as prototype theories in categorization. All current models collapse the many contexts that a word occurs in to a single best-fitting representation, but that process discards idiosyncratic regularities that are important to word meaning. Homonymous words present an ideal evaluation case. It has long been known that spatial DSMs collapse the multiple senses of a homonym into a single representation, often averaging over very distinct context patterns to a center of tendency that represents the average meaning. A word such as *bank* will be positioned in space as a frequency weighted average of its distinct senses.

Griffiths, Steyvers, and Tenenbaum (2007) have suggested that homonyms and polysemes pose a core challenge to spatial DSMs that they cannot adequately explain, arguing instead for probabilistic topic models. In addition, homonyms and polysemes are hardly rare in language: Over half of all words in English have multiple senses, and the frequency distribution of senses for a word tends to be positively skewed. DSMs lose the tail when collapsing to a prototype, but humans can regularly comprehend the multiple (less frequent) meanings that are averaged out in DSMs. Hence, DSMs have great difficulty with the non-dominant sense of homonyms (e.g., the river sense of *bank* is dominated by the financial institution sense in the prototype representation). Homonyms may be a key falsification criterion for DSMs that posit abstraction at learning.

In this paper we present a different notion of abstraction. Building on successful instance-based memory models, we

posit that semantic abstraction may be a consequence of retrieval from episodic memory rather than a learning mechanism. We present an instance-based theory of semantics (ITS) that stores word contexts as multiple instances in episodic memory. When a word is presented to the model, a simple retrieval mechanism is applied that generates an ad hoc semantic representation of the word. In contrast to abstraction-at-learning DSMs, ITS is able to have non-linear activation of encoded instances, which allows it to easily access the non-dominant sense of a word when provided the appropriate context.

Instance-Based Theory of Semantics

ITS is rooted in Hintzman's (1986) MINERVA 2 instance-based model of human memory (see Johns & Jones, 2015, and Kwantes, 2005, for related approaches).

In the theory, every letter string (i.e., word or non-word) is represented by a unique n dimensional vector, w, where each dimension takes a randomly sampled value from a normal distribution with mean zero and standard deviation $1/\sqrt{n}$. Vectors constructed in this manner are orthonormal in expectation and, thus, the model begins from a state in which words have no similarity to one another.

Memory of a conversation, a document, is encoded as an instance d_i, equal to the sum of the $j = 1 \ldots h$ words in the document,

$$d_i = \sum_{j=1}^{j=h} w_{ij}$$

where h is the number of words in document i, w_j is word j in the document, and d_i is the sum of the words in document i. To illustrate, the document, "the dog bit the mailman" is stored as $w_{dog} + w_{bit} + w_{mailman}$ (consistent with standard practice, we excluded a list of stop words).

Memory proper is a collection of document representations in which each document i, d_i, is stored to a corresponding row in a memory matrix, M,

$$M_i = d_i = \sum_{j=1}^{j=h} w_{ij}$$

To retrieve a word's meaning, a word vector is presented to memory as a probe and a corresponding semantic vector, c, is retrieved that is called the *echo*.

Retrieving the echo is a two-step process. In step one, the probe, p, composed of h words activates all traces in memory, M, in parallel,

$$a_i = \prod_{k=1}^{k=h} \left(\frac{\sum_{j=1}^{j=n} p_{kj} \times M_{ij}}{\sqrt{\sum_{j=1}^{j=n} p_{kj}^2} \sqrt{\sum_{j=1}^{j=n} M_{ij}^2}} \right)^3$$

where a_i is the activation of trace i, p_{kj} is feature j of word k in the probe, M_{ij} is feature j of document i in memory, n is the dimensionality of a word representations, and h is the number of words in the probe. Cubing the cosine similarity between probe and trace forces a more selective retrieval of traces that are most similar to the probe. Activation ranges between -1 and +1: when the trace and probe are identical $a = 1$, when the trace and probe are orthogonal $a = 0$, and when the trace and probe are opposite $a = -1$.

As should be obvious, the product rule in the activation function supports a selective activation of traces that include all h words in the probe more strongly than traces that include a subset of the h words in the probe. This feature of the model will be critical for supporting contextual disambiguation of word meaning.

In step two, an aggregate of the activated traces is retrieved that is a vector called the *echo*. The contribution of each trace to the echo is in proportion to its activation,

$$c_j = \sum_{i=1}^{i=m} \sum_{j=1}^{j=n} a_i \times M_{ij}$$

where c_j is feature j in the echo, m is the number of traces in memory, a_i is the activation of trace i, and M_{ij} is the value of feature j in trace i in memory. The echo is the corresponding semantic representation retrieved for the probe.

Finally, the semantic resemblance, r, between two probes (e.g., two words), p_1 and p_2, is computed as the cosine similarity between their corresponding echoes,

$$r(p_1, p_2) = \frac{\sum_{j=1}^{j=n} c_{1j} \times c_{2j}}{\sqrt{\sum_{j=1}^{j=n} c_{1j}^2} \sqrt{\sum_{j=1}^{j=n} c_{2j}^2}}$$

where c_1 is the echo retrieved by p_1 and c_2 is the echo retrieved by p_2. Thus, words that retrieve similar echoes are judged similar in meaning.

In summary, the theory assumes that people remember their language experiences and that word meaning is constructed during retrieval. We now turn to a demonstration of the theory using a simple artificial language.

Artificial Language Simulations

Natural language is a complex structure and, therefore, is hard to assess cleanly. To finesse the problem, we developed a small artificial language.

Our toy language is presented in Figure 1. It was inspired by languages from Lee (1962) and Elman (1990).

Our toy language included seven different word classes, with each word class represented by two words. For example, the class NOUN_HUMAN was represented by the words *man* and *woman* whereas the class VERB_VEHICLE was represented by *stop* and *break*.

The language also included three sentence frames that can be re-written as sentences permissible in the language.

A permissible sentence is generated by selecting a sentence frame and then rewriting the word classes with words from that class. For example, the template NOUN_HUMAN, VERB_DINNERWARE, NOUN_DINNERWARE can produce "*man smash plate*", "*man break plate*", "*man smash glass*", "*man break glass*", "*woman smash plate*", "*woman break plate*", "*woman smash glass*", and "*woman break glass*" by applying the following rewrite rules: (a) NOUN_HUMAN → {*man*, *woman*}, (b) VERB_DINNERWARE → {*smash, break*}, and NOUN_DINNERWARE → {*plate, glass*}. The full list of 24 sentences (8 per sentence frame) defines the language.

Categories of lexical items:

Categories	Examples
NOUN_HUMAN	*man*, *woman*
NOUN_VEHICLE	*car*, *truck*
NOUN_DINNERWARE	*plate*, *glass*
NOUN_NEWS	*story*, *news*
VERB_VEHICLE	*stop*, ***break***
VERB_DINNERWARE	*smash*, ***break***
VERB_NEWS	*report*, ***break***

Sentence frames:

NOUN_HUMAN; VERB_VEHICLE; NOUN_VEHICLE
NOUN_HUMAN; VERB_DINNERWARE; NOUN_DINNERWARE
NOUN_HUMAN; VERB_NEWS; NOUN_NEWS

Figure 1: The toy language.

Critical for our analysis, the word *break* is included as a word in all three verb classes. This makes *break* a homonym with three different senses. In the vehicle sense, it is related to *stop*. In the dinnerware sense, it is related to *smash*. In the news sense, it is related to *report*.

Simulations with balanced sentence frequency

We conducted simulations with ITS for a corpus of sentences from our toy language. In a first set of simulations, every sentence was equiprobable in the corpus.

For each simulation, we generated a corpus of 20,000 sentences, where each sentence was sampled with equal probability. Then, we applied our model to retrieve a semantic vector (i.e., the echo) for (a) each individual word in the language (i.e., *man, woman, car truck, plate, glass, news, story, stop, smash, report, break*) and (b) each pair of words in the language (e.g., *man*/*car*, *man*/*truck*, and so on).

The single-word similarities are summarized in the top left panel in Figure 2, with the semantic structure of word meaning drawn as a two-dimensional MDS plot (Shepard, 1980).

As shown, ITS captured the structure of word meaning. Firstly, words belonging to the same topic are clustered together. Secondly, words belonging to different topics are separated. Thirdly, the homonym (i.e., *break*) is equidistant to the vehicle, dinnerware, and news clusters.

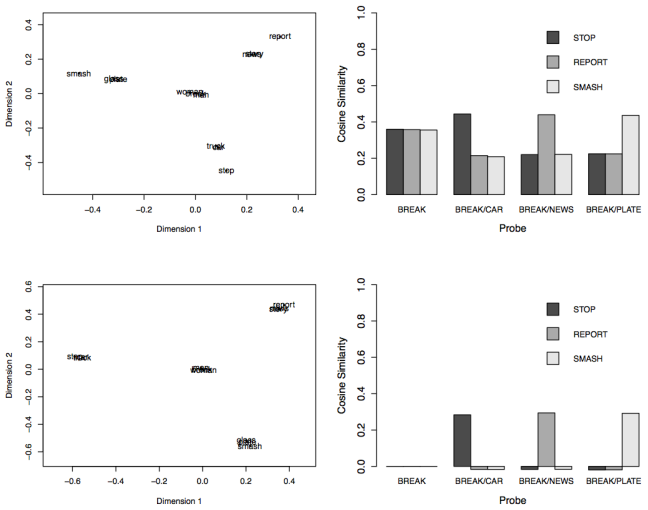

Figure 2. Toy language simulations with balanced sentence frequency. Results with ITS are presented in the top row. Results with LSA are presented in the bottom row.

The top right panel of Figure 2 shows ITS's ability to disambiguate the meaning of *break* depending on the context in which it is presented. As shown, presenting ITS with *break* in isolation retrieves an echo that is equally similar to all three of its potential meanings (i.e., *stop*, *smash*, and *report*). However, presenting *break* in conjunction with *car* retrieves an echo that is more similar to *stop* than to either *smash* or *report*, presenting *break* in conjunction with *story* retrieves an echo that is most similar to *report,* and presenting *break* in conjunction with *plate* retrieves an echo that is most similar to *smash*.

For comparison, we conducted corresponding simulations with LSA (Landauer & Dumais, 1997). In those simulations, we derived the word-by-document matrix from the same corpus, weighted the matrix by the standard entropy calculation, derived a solution by dimension reduction, computed the cosine similarity between words in each of the reduced spaces, and computed the similarity between the word vectors. Two-word probes were presented as the sum of the corresponding vectors.

The bottom row in Figure 2 presents the corresponding results with LSA. As shown, LSA and ITS arrive to very similar solutions.

In summary, when sentence frequency is balanced, both ITS and LSA (a) recover the structure of a small artificial language, (b) recognize homonymy, and (c) disambiguate the intended meaning of a contextually-signaled homonymous word (i.e., *break*).

Simulations with unbalanced sentence frequency

Our toy language differs from natural language in important ways. For example, the homonymous word *break* has no dominant sense: it is as likely to mean *stop*, *smash,* or *report*. In this simulation, we evaluate both ITS and LSA against a corpus constructed so that *break* has a dominant

sense.

To give *break* a dominant sense, we constructed a new corpus composed of more sentences from the vehicle topic (p = 4/6) than from the dinnerware (p = 1/6) and news (p = 1/6) topics. Thus, *break* appeared more often in the *stop* sense than the *smash* and *report* senses. Otherwise, the simulation was identical to the one already presented.

Results are presented in Figure 3. ITS's behaviour is shown in the top row; LSA's in the bottom row.

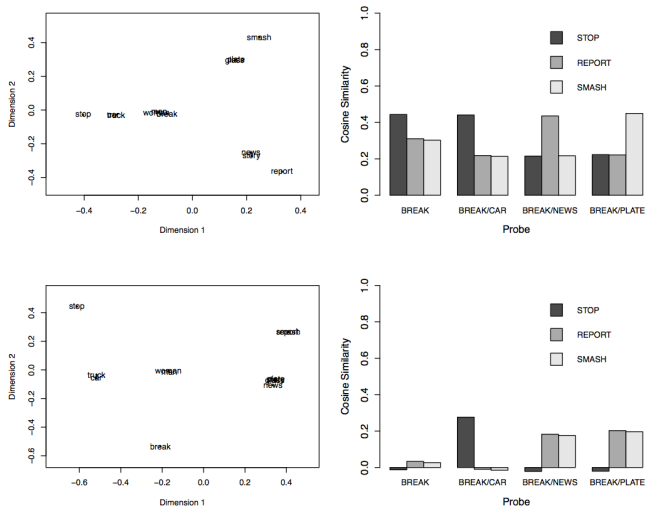

Figure 3. Toy language simulations with unbalanced sentence frequency. Results with ITS are presented in the top row. Results with LSA are presented in the bottom row.

As shown, ITS's behaviour was affected by the manipulation, but in sensible ways. The model recognizes that *break* presented in isolation has a dominant sense; but, it also retrieves the contextually appropriate sense of *break* when presented in conjunction with a disambiguating noun (e.g., *break* is more similar to *smash* when presented in conjunction with *plate*). In contrast, LSA's behavior was strongly and adversely affected by the manipulation. The meaning of *break* presented in isolation contradicts the word's dominant sense (i.e., *break* is more similar to *smash* and *report* than it is to *stop*). More importantly, the model fails to disambiguate between the subordinate meanings of *break* when it is presented in combination with a noun associated with one of its subordinate meanings. The demonstration confirms that a prototype model of semantics fails to disambiguate word meaning and, thus, offer a compromised descriptive account of semantic knowledge. More central to our argument, it also shows that an instance-based approach to semantics solves the problem.

Natural Language Simulations

The simulations presented so far give a good picture of our instance-based model of semantics and how it disambiguates the meaning of a homonym presented in context. However, solving a toy problem does not guarantee a solution to the problem at scale. Thus, we applied ITS at scale to a record of natural language experience.

Taxonomic structure

A benchmark requirement of semantic theories is that they can organize words into coherent taxonomic categories. For example, a competent theory of semantics should recognize that items from the category of *animals* are more similar to one another than they are to items from the category of *vehicles*.

We evaluated ITS against that criterion by storing a record of language experience from the Touchstone Applied Science Associates (TASA) corpus and retrieving echoes for words from the well-defined taxonomic categories used in previous work with BEAGLE (Jones & Mewhort, 2007, Figure 3) and HAL (Lund & Burgess, 1996, Figure 2).

The top row in Figure 4 presents ITS's organization of words from three taxonomic categories (i.e., finance, science, and sports) on the left and its organization of words from three other taxonomic categories (i.e., *animal names*, *body parts*, and *geographic locations*) on the right.

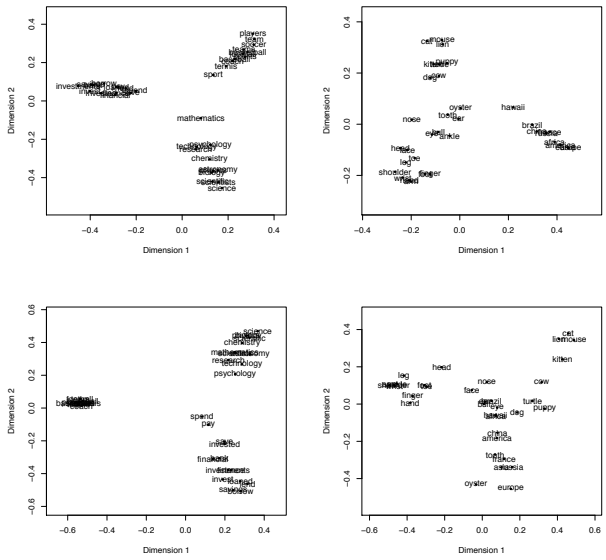

Figure 4. Taxonomic categories. Results with ITS are presented in the top row. Results with LSA are presented in the bottom row.

As shown, ITS does an excellent job of grouping words in the same category and distinguishing words from different categories.

To confirm the visual impressions given by the MDS solutions, we computed the intracategory and intercategory similarities between words. For the Jones and Mewhort graph, the mean intracategory item-to-item cosine similarity (M = .27, SD = .11) was, by a conservative estimate, 1.82 standard deviations greater than the mean intercategory

item-to-item similarity (M = .07, SD = .04). The same is true for the Lund and Burgess graph: the mean intracategory item-to-item cosine similarity (M = .18, SD = .10) was, by a conservative estimate, a still strong 1 standard deviations greater than the mean intercategory item-to-item similarity (M = .08, SD = .05).

Results with LSA are presented in the bottom row of Figure 4. Although LSA accomplished the discriminations, it did not perform as well as ITS. For the Jones and Mewhort (2007) set, the mean intracategory item-to-item cosine similarity (M = .42, SD = .28) was 1.50 standard deviations greater than the mean intercategory item-to-item similarity (M = .00, SD = .03). For the Lund and Burgess (1996) set, the mean intracategory item-to-item cosine similarity (M = .14, SD = .19) was 0.68 standard deviations greater than the mean intercategory item-to-item similarity (M = .01, SD = .05).

In summary, the results serve proof of concept that an instance model of semantics can perform taxonomic classification. The results also show that it can perform the discrimination in the same quantitative range as an established prototype model of semantics.

Disambiguation of word meaning

ITS can group words that have related meanings, but that doesn't mean that it can disambiguate the meaning of a homonym conditional on context. To evaluate the problem, we applied ITS to the disambiguation of homonyms from a lexical decision study by Schvaneveldt, Meyer, and Becker (1976).

On each trial in the experiment, participants were presented with three successive letter-strings (e.g., *save-bank-money* or *save-bank-boat*) and required to identify each one as a word or nonword. On *cued trials*, the first two strings cued the appropriate meaning of the third string (e.g., *river*/*bank* cued *boat*). On *miscued trials*, the first two words miscued the appropriate meaning (e.g., *river*/*bank* miscued *money*). The critical result (or at least the one relevant here) was that people were faster to identify the third word on cued compared to miscued trials.

To evaluate ITS, we conducted a simulation using Schvaneveldt et al.'s (1976) materials (see their Table 2, p. 248). On each trial, an echo was retrieved for the joint probe composed of the first and second words (e.g., *river*/*bank*), an echo was retrieved for the third word (e.g., *money*), and the two echoes were compared.

If ITS solves the problem, the similarity between the echo retrieved by words one and two and the echo retrieved by word three will be greater on cued than miscued trials.

We conducted a full set of 432 comparisons to match the original experiment: all 144 cued trials and 288 miscued trials. To summarize performance, we computed a mean and variance for the cosines from all 144 of the cued trials and from all 244 of the miscued trials. We also computed corresponding simulations with LSA.

Results are presented in the left column of Figure 5; whiskers show the standard error of the mean.

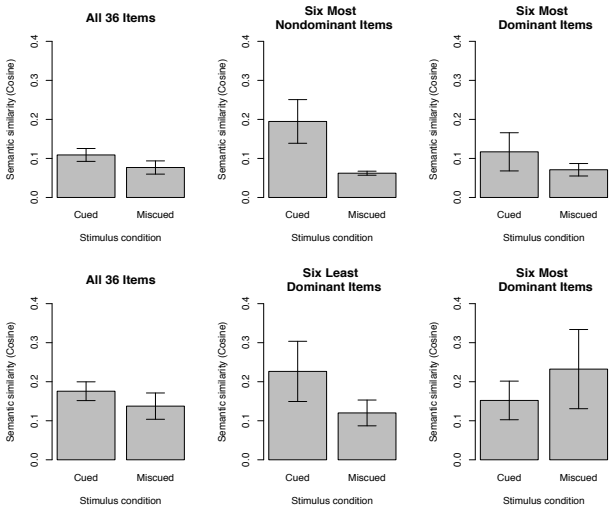

Figure 5. Disambiguation of homonyms. Results with ITS are presented in the top row. Results with LSA are presented in the bottom row.

As shown in Figure 5, both ITS and LSA anticipate the cued versus miscued difference, with the mean similarity of echoes retrieved by the primes (i.e., words one and two in conjunction) and probes (i.e., the third word) greater for cued than miscued trials.

At first blush, the results suggest that both a prototype and instance-based model can disambiguate the meaning of a homonymous word. But, the analysis does not distinguish performance depending on whether a homonym does or does not have a dominant meaning.

To examine that problem, we used empirical norms from Armstrong, Tokowicz, and Plaut (2012) to identify items in Schvaneveldt et al.'s (1976) stimulus set that do and do not have dominant meanings. Then, we re-calculated performance as a function of that distinction.

The centre and right columns in Figure 5 show results for the six most and six least dominant homonyms, respectively. Consistent with our earlier analysis using the toy language, ITS succeeded at understanding the cued meaning of both a dominant and subdominant homonym. But, LSA did not. In fact, LSA produced a qualitatively different pattern that indicates an outright failure to retrieve the subordinate word sense.

Discussion

Prototype models of semantics represent a sophisticated leap forward for understanding the acquisition and representation of word meaning. However, they have difficulty understanding the intended meaning of ambiguous words, and this may signal an architectural flaw with the notion of abstraction at learning.

To test the notion of abstraction at retrieval, we developed an instance-based approach to the retrieval of knowledge. In contrast to prototype theories that encode word meaning prospectively, our approach assumes that

word meaning is constructed retrospectively, as a consequence of retrieval from a decentralized and episodic record of language experience. Our simulations join with a small body of positive evidence for an instance-based approach to semantics (Johns et al., 2015; Kwantes, 2005). However, it joins a large body of positive evidence that instance theories outperform prototype theories in the domain of knowledge and categorization.

One reason for the predominance of prototype-based theories may be due in part to Chomskian presumptions in linguistics: that the job of the cognitive mechanism is to induct and abstract the rules of a grammar from instances. This abstractionist presumption may have implicitly guided architectural decisions in preceding prototype accounts, even as work in related domains (e.g., artificial grammar learning) has accumulated evidence in favour of an instance- over prototype-based explanation of behavior (e.g., Jamieson & Mewhort, 2009).

A second reason is cognitive economy. When Rosch and Mervis (1975) developed their prototype and hierarchical methods for knowledge representation, a guiding principle for semantics was cognitive economy. However, our instance-based approach coupled with recent work on usage-based theories of linguistics force a reconsideration of economy as a forcible constraint on theory development (e.g., Tomasello, 2003).

A third reason is that the computational economy of the prototype approach fits better with the speed/accuracy tradeoff for application development. In the prototype method, the vectors are derived and then applied consistently thereafter. In the instance method, semantic vectors must be retrieved on-the-fly and, thus, requires a continuous derivation of semantics. For researchers developing search engines, the speed of retrieval matters and, so, the benefits of instance-based approach might be outweighed by the need to present a record of documents quickly to the user. Moving forward, we will consider the ITS's instance-based solution to word-sense disambiguation against the solutions developed in extended prototype accounts that encode multiple prototypes to encode the different senses of a word (e.g., Erk & Pado, 2008; Reisinger & Mooney, 2010).

In some ways, it is tempting to see instance-based DSMs as "cheating". If the model stores all data, then it can compute an accurate semantic representation whenever one is needed. But the theoretical claim is profound in its proposal: we may not have semantic memory in the way that theorists have typically conceived of semantic memory.

In place of the standard view, an instance-based approach to semantics proposes that a person's interpretation of the words they are reading is constructed during retrieval and on-the-fly such that our phenomenology of meaning is continuously constructed by the interaction of stimuli and experience (Kintsch & Mangalath, 2011). But the instance-based approach should also put us at ease because it provides converging evidence that performance across multiple cognitive domains (e.g., categorization, memory, semantics) might be explicable from the same core cognitive principles.

Acknowledgments

This research was supported by a NSERC grant to RKJ. We want to thank the University of Manitoba and Indiana University for sabbatical support to RKJ where this work was completed.

References

Armstrong, B. C., Tokowicz, N., & Plaut, D. C. (2012). eDom: Norming software and relative meaning frequencies for 544 English homonyms. *Behavior Research Methods*, *44*, 1015-1027.

Erk, K., & Padó, S. (2008, October). A structured vector space model for word meaning in context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 897-906). Association for Computational Linguistics.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological review*, *114*, 211.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Jamieson, R. K., & Mewhort, D. J. K. (2009). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology*, *62*, 550-575.

Johns, B. T., & Jones, M. N. (2015). Generating structure from experience: A retrieval-based model of language processing. *Canadian Journal of Experimental Psychology*, *69*, 233-251.

Jones, M. N. (2018). When does abstraction occur in semantic memory: Insights from distributional models. *Language, Cognition and Neuroscience*.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review, 114,* 1–37.

Jones, M. N., Willits, J. A., & Dennis, S. (2015). Models of semantic memory. In J. R. Busemeyer & J. T. Townsend (Eds.) *Oxford Handbook of Mathematical and Computational Psychology*. 232-254.

Kintsch, W. (2001). Predication. *Cognitive science*, *25*, 173-202.

Kintsch, W., & Mangalath, P. (2011). The construction of meaning. *Topics in Cognitive Science*, *3*, 346-370.

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review, 12,* 703–710.

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

Lee, S. (1962). *Incredible hulk*. Marvel Comics.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*, 203-208.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

Reisinger, J., & Mooney, R. J. (2010, June). Multi-prototype vector-space models of word meaning. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 109-117). Association for Computational Linguistics.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*, 390-398.

Schvaneveldt, R. W., Meyer, D. E., & Becker, C. A. (1976). Lexical ambiguity, semantic content, and visual word recognition. *JEP:HPP*, *2*, 243-256.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.