

Construction in Semantic Memory: Generating Perceptual Representations With Global Lexical Similarity

Brendan T. Johns (johns4@indiana.edu)

Michael N. Jones (jonesmn@indiana.edu)

Department of Psychological and Brain Sciences, Indiana University
1101 E. Tenth Street, Bloomington, In., 47405

Abstract

The literature currently contains a dichotomy in explaining how humans learn lexical semantic representations for words. Theories generally propose either that lexical semantics are learned through perceptual experience, or through exposure to regularities in language. We propose here a model to integrate these two information sources. The model uses the global structure of memory to exploit the redundancy between language and perception in order to generate perceptual representations for words with which the model has no perceptual experience. We test the model on a variety of different datasets from embodied cognition experiments.

Keywords: Semantic memory; co-occurrence models; LSA.

Introduction

Modern computational models of lexical semantics (e.g., latent semantic analysis (LSA); Landauer & Dumais, 1997) infer representations for words by observing distributional regularities across a large corpus of text. Although their specific learning mechanisms may differ considerably, all members of this class of model rely on statistical information in text to infer semantic structure. Distributional models have seen considerable success at accounting for an impressive array of behavioral data in tasks involving semantic cognition. Since their beginning, however, distributional models have been heavily criticized for their exclusive reliance on linguistic information (e.g., Perfetti, 1998), essentially making them models of learning meaning “by listening to the radio” (McClelland).

More recently, empirical research has demonstrated that distributional models fail to account for a variety of semantic phenomena in the realm of embodied cognition (e.g., Glenberg & Robertson, 2000). This failure is not a great surprise given that distributional models do not receive perceptual input, and they actually perform surprisingly well on many tasks believed to require perceptual learning due to the amount of perceptual information redundantly coded in both language and the environment (for a review, see Riordan & Jones, 2010). Distributional models do not argue that perceptual information is unimportant to semantic learning. Perceptual information is still statistical information; what is required is a mechanism by which these two sources of information may be integrated. Attempts to integrate linguistic and perceptual information in a unified distributional model are now emerging (e.g., Andrews, Vigliocco, & Vinson, 2009; Jones & Recchia, 2010). However, there is little connection in these models to existing theories of modal perceptual symbol learning.

Perceptual symbol systems theory (PSS; Barsalou, 1999), one of the cornerstones of the grounded cognition movement (Barsalou, 2008), has been proposed as a competitor to distributional models as an explanatory theory for the emergence of lexical semantic structure in memory. The basis of PSS is the dismissal of amodal symbols as the central component underlying human mental representation. Rather, the PSS approach proposes that the symbols used in reasoning, memory, language, and learning are grounded in sensory modalities.

In the realm of lexical semantics, PSS proposes that the mental representation of a word is based on the perceptual states that underlie experiences with the word’s physical referent (Barsalou, 1999). Across many experiences with words, the underlying neural states tend to stabilize and create an accurate perceptual representation of a word that is grounded across sensory areas in the cortex. There is considerable evidence, across both behavioral and neuroimaging experiments, that the perceptual associates of words do play a central role in language processing (for a review see Barsalou, 2008).

Although distributional models and PSS are often discussed as competing theories, the two are certainly not mutually exclusive. PSS is unable to make claims about the meanings of words that have no physical manifestation—it is fairly limited to concrete nouns and action verbs (although these are the most commonly used experimental stimuli). Further, PSS is silent regarding the simple observation that humans are quite capable of forming sophisticated lexical representations when they have been given nothing to ground those representations in. This is the situation in which distributional models excel—inferring the meaning of words in the absence of perceptual information. However, distributional models certainly fail when given tests that stress the use of perceptual information—the situation in which PSS excels. Hence, the two theories should not be viewed as competitors, but rather as complimentary (see Riordan & Jones, 2010). What is needed is research into how humans might integrate the two types of information to make full use of both the structure of language and the perceptual environment.

Here we explore whether a central component of PSS, *perceptual simulation*, may be integrated with a distributional model to infer perceptual information for words that have never been “perceived” by the model based on global lexical similarity to words that have been perceived. Further, we test the model’s ability to infer the

likely linguistic distributional structure for a word in absence of linguistic experience from its perceptual similarity to words with which the model has had linguistic experience. In this sense, the model’s goals are similar to previous integrative attempts (Andrews et al., 2009; Jones & Recchia, 2010), but is theoretically linked to important mechanisms in PSS.

PSS proposes that simulations (based on past experiences) play a central role in conceptual and semantic processing, and there is a considerable amount of evidence that this is a mechanism of central importance in human cognition (Barsalou, et al., 2003). PSS presumes that each lexical representation is a multi-modal simulation of the perceptual experience of that word (e.g. the simulator for horse may contain what a horse looks like, feels like, sounds like, how you ride one, etc...), which is reinstated whenever one experiences a word. For example, when reading the word *metal*, your semantic representation is a simulation of previous perceptual experiences with the word’s referent, including its texture, experiences of hard and cold, etc. The word’s meaning is not disembodied from its perceptual characteristics.

We by no means have a solution as to how to formalize this simulation process, but instead evaluate a type of simulation that may underlie our ability to make correct inferences about the perceptual representation of ungrounded words. Instead of relying upon the structure of neural states during experience, it instead relies upon the grounded representations of other words. That is, a word’s perceptual simulator can be constructed not by the current perceptual state, but by the perceptual states of similar words in memory. The importance of a given word’s state is determined by the associative strength between the two words, derived from the statistical structure of how those words are used in the language environment. Hence, global lexical similarity (similarity of a word to all other words in memory) may be used by a generation mechanism to ‘fill-in’ the perceptual representation for a specific word. We integrate this idea of experiential simulation into a global memory model of semantics, based loosely on Hintzman’s (1986) MINERVA 2 model.

Generating Perceptual Representations

It is important that we are clear at the outset in our definitions of linguistic, perceptual, and lexical information in this model, as they are clearly oversimplifications. A word’s *linguistic* information in the model is very simply a vector representing its co-occurrence structure across documents in a text corpus. If the word is present in a given document, that vector element is coded as one; if it is absent, it is coded as zero. A word’s *perceptual* information in the model is a probability vector over perceptual features generated by human subjects. For example, the feature *<has_fur>* will have a high probability for *dog*, but a low probability for *pig*, and a zero probability for *airplane*. It is important to note that these types of feature norms include much information that is non-perceptual (e.g., taxonomic,

situational), and are unable to represent more complex perceptual information such as embodied interaction; nonetheless, they are a useful starting point. A word’s full *lexical* representation in the model is simply the concatenation of its linguistic and perceptual vectors (even if one of the two is completely empty). We demonstrate that this model is able to use a simple perceptual simulation mechanism to account for a diverse set of both behavioral and neuroimaging results in studies of language processing.

Linguistic co-occurrence vectors for words were computed from counts across 250,000 documents extracted from Wikipedia (Recchia & Jones, 2009). Perceptual vectors will depend on the particular simulation, but will include feature generation norms (McRae, Cree, Seidenberg, & McNorgan, 2005; Vinson & Vigliocco, 2008), and modality exclusivity norms (Lynott & Connell, 2009). Each word’s representation in the full memory matrix is a concatenation of its linguistic and perceptual vectors. The goal of the model is to infer the perceptual vector for a word from global linguistic similarity to other words, using this limited data to generalize to the entire lexicon.

Borrowing from Hintzman’s MINERVA model (see also Kwantes, 2005), our model attempts to create an abstraction of a word’s full lexical vector using a simple retrieval mechanism. When a partial probe is compared to memory (say, a word with a linguistic vector, but a zero perceptual vector), a composite ‘echo’ vector is returned consisting of the sum of all lexical vectors in memory weighted by their similarity to the probe. Across the lexicon, this returns a stable full lexical estimate for a word, including an inferred perceptual vector. Specifically, perceptual representations are constructed in a two-step abstraction process, based on Hintzman’s process of ‘deblurring’ the echo.

In **step 1** each representation in memory with a zero perceptual vector has an estimated perceptual vector constructed based on its weighted similarity to lexical entries that have non-zero perceptual vectors:

$$Perc_j = \sum_{i=1}^M T_i * S(T_i, T_j)^\lambda, \quad (1)$$

Where M represents the size of the lexicon, T represents the lexical trace for a word, S is similarity function (here, vector cosine), and λ is a similarity weighting parameter. Lambda is typically set to 3 (Hintzman, 1986), but we will fit this parameter for each of the different norms (due to differences in their dimensionality and structural characteristics), and also for the two different steps of inference (due to differences in the number of traces being used to create an echo). Step 1 utilizes only a limited number of traces and so each trace should add more information, while in Step 2 the entire lexicon is used, and so each word trace should be more limited in its importance.

In **step 2**, the process from step 1 is iterated, but inference for each word is made from global similarity to *all* lexical entries (as they all now contain an inferred perceptual vector). Hence, representations in step 1 are inferred from a limited amount of data (only words that have been “perceived” by the model). In step 2, representations for

each word are inferred from the full lexicon—aggregate linguistic and perceptual information inferred from step 1.

This two-step process is illustrated in Figure 1. Prior to the inference process, only linguistic information is contained in memory with a limited amount of perceptual information. Across the two-step abstraction process, the model is able to use the associative structure of memory, along with this initially limited amount of data, and infer a perceptual representation for each word. The essential claim of this model is that the global similarity structure that is contained in the lexicon is sufficient to make sophisticated predictions about the perceptual properties of words.

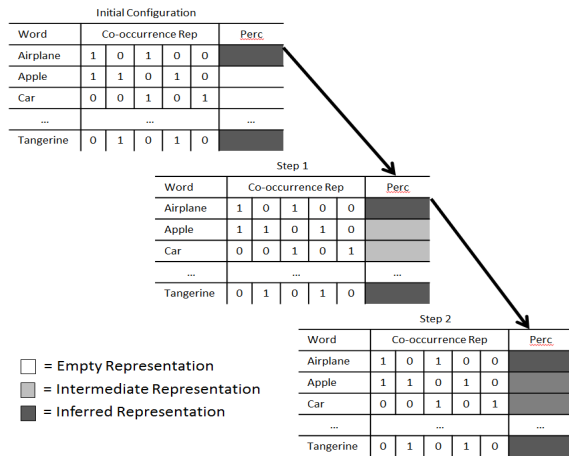


Figure 1. The two-step process of global construction.

1. Testing Model Foundations

Our preliminary examination of this model will be done by manipulating core aspects of the framework, including training the model with different perceptual norms, changing the lexicon size, and testing on different corpora.

Simulation 1.1: Word Norms

Two different types of perceptual norms were used for evaluation: feature generation norms (McRae, et al., 2005; Vinson & Vigglioco, 2008) and modality exclusivity norms (Lynnott & Connell, 2009). Feature generation norms are created from hundreds of subjects producing the perceptual features for a set of target words. Aggregated across subjects, the result is a vector across possible features for each word, with elements representing the generation probability of a given feature for a given word. Modality exclusivity norms are created by having subjects rate how much a target word is based in each of the five sensory modalities. The result is a five-element vector per word, with each element representing the strength of that modality for a given word.

To evaluate how well the model is able to infer a word’s perceptual representation, we used a cross-validation procedure. For each sample, a word was randomly selected from the perceptual norm of interest, and its perceptual vector in the lexicon was zeroed out. The model then infers

a perceptual representation for the blanked out word based on its associative similarity to other words in the lexicon across our two inference steps. Finally, the correlation is computed between the inferred perceptual vector and the true perceptual vector in the norms for the target word. This procedure was conducted across all words in each of the norms. For perceptual norm and step, the λ parameter was hand fit to the data.

Table 1. Model Predictions for each Word Norm

| Word Norm | Step 1 | Step 2 |
|--------------------|--------|--------|
| McRae, et al. | 0.42 | 0.72 |
| Vinson & Vigglioco | 0.42 | 0.77 |
| Lynnott & Connell | 0.83 | 0.85 |

* All correlations significant at $p < 0.001$

The correlations for each of the word norms across the two steps are displayed in Table 1. This table shows that for each of the norms that model is able to infer an accurate perceptual representation is at a high level, with all three norms achieving a correlation above 0.7.

Simulation 1.2: Effect of Lexicon Size

A second simulation was conducted to manipulate the number of words in the lexicon used to create the inferred perceptual representations. This was done by varying the number of words in the lexicon from 2,000 \rightarrow 24,000 in steps of 2,000. The lexicon was arranged by frequency from the TASA corpus such that only the most frequent set of words are included. This simulation exclusively used the norms from McRae, et al. (2005).

The magnitude of correlation as a function of lexicon size is shown in Figure 2. This figure shows that a consistent increase in fit is attained as the size of the lexicon grows, until about a size of 14,000. From that point on, the model produces a reduced fit. The reason for this is that after 14,000 words the amount of noise that is accumulated within the echo vector exceeds the benefits of the added resolution created by the additional associative structure provided by the increased lexicon size. In the following simulations only the first 14,000 words will be utilized by the generation mechanism.

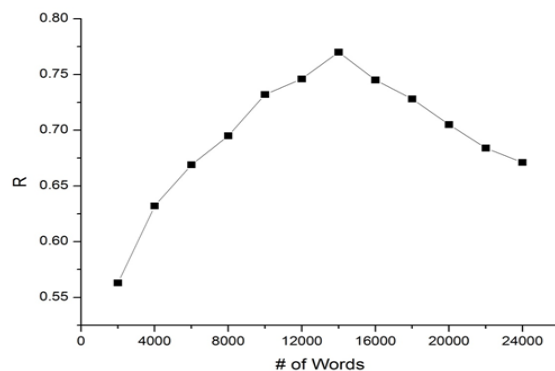


Figure 2. Effect of lexicon size

Simulation 1.3: Effect of Corpus Size

Recchia & Jones (2009) have demonstrated that increasing the size of a corpus (i.e. increasing the number and diversity of the contexts that a word appears in) also increases the fit to semantic similarity ratings, independent of abstraction algorithm. To evaluate this trend for inferring perceptual representations in our global similarity model, we compared the goodness-of-fit for the model predictions of the McRae, et al. (2005) norms over a small corpus (the TASA corpus, composed of 37,600 documents) and a large corpus (a Wikipedia corpus, composed of 250,000 documents). The fit for the TASA corpus was $r = 0.34$ after the first step, and $r = 0.64$ after the second step. However, with the larger Wikipedia corpus, a correlation of $r = 0.42$ after the first step, and an $r = 0.77$ after the second step. This shows that there is an impressive increase in fit between the model’s predictions and data with the use of a larger corpus, even though the TASA corpus is of higher quality. This is an important result: It demonstrates that the greater the amount of experience the model has with language, the better its inferences are about a word’s perceptual representation.

Simulation 1.4: Reverse Inference

An interesting aspect of this model is that it is capable of making reverse inferences. Given the perceptual representations for words, the model should be able to estimate the likely linguistic distributional structure for a word. To test reverse inference in the model, we estimated each word’s perceptual representation using (1). A word’s inferred linguistic vector was then estimated with (1), but rather than summing across the perceptual representations in the lexicon, the linguistic vectors were used (and similarity was based on similarity of perceptual vectors). The inferred linguistic vector was then correlated with the word’s retrieved co-occurrence vector, where the probe vector is co-occurrence representation is word, and the representation of other words is summed, similar to Kwantes (2005).

The correlation between the inferred linguistic representations for the concrete nouns from the McRae, et al. norms was $r = 0.67$, $p < 0.001$. For all other words in the lexicon, this correlation is $r = 0.5$, $p < 0.001$. The second set is lower than the concrete nouns for two reasons: 1) the perceptual space of the McRae norms does not extend to all words, and 2) not all words have a strong perceptual basis (e.g. abstract words) and so the inferred perceptual vector not diagnostic of that word’s meaning. However, this simple analysis does show that the model is capable of this reverse inference: it can, given the perceptual representation of a word, construct a fairly accurate approximation of the linguistic co-occurrence structure of that word.

This is a central finding for the model because it allows for lexical inferences to be made in two directions, both from linguistic to perceptual and from perceptual to linguistic. Hence, the model can take in either perceptual or linguistic information about a word and infer the other type of representation from it, allowing for both aspects of memory to be filled in when information is missing.

2. Behavioral Simulations

The set of simulations in this section uses the global similarity model to evaluate the model’s predictions of a variety of behavioral phenomena from grounded cognition.

Simulation 2.1: Affordances

In a test of the strength of distributional models (specifically, LSA) Glenberg & Robertson (2000) conducted a study in which they assessed subjects’ (and LSA’s) ability to account for affordance ratings to different objects within a given sentence. Objects ranged from being realistic within the context of the sentence, to being afforded, or non-afforded. For example, subjects were given the sentence “Hang the coat on the _____”, and were asked to give ratings on three words (realistic = *coat rack*, afforded = *vacuum cleaner*, and non-afforded = *cup*). Unsurprisingly, realistic objects had a higher score than both afforded and non-afforded objects, and afforded objects had a higher rating than non-afforded objects. However, the stimuli were constructed such that LSA could not discriminate between afforded and non-afforded conditions.

Our model is not a model of sentence comprehension (neither is LSA), so a simpler test was conducted using Glenberg and Robertson’s (2000) stimuli. The central action word that described the affordance was used (e.g. “hang” instead of “Hang the coat on the _____”). Then the cosine between this target word and the three different object words were calculated for both the inferred feature vectors and the raw co-occurrence vectors. The norms from McRae et al. (2005) were used for this test. The results of this simulation are displayed in Figure 3.

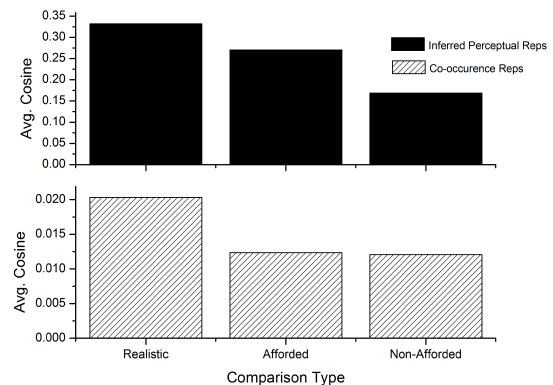


Figure 3. Simulation of results using stimuli from Glenberg & Robertson (2000).

As shown in Figure 3, the inferred feature vectors are able to generate the correct pattern of results – that is, the average cosine for the realistic words is greater than for the afforded and non-afforded words, and also the average cosine for the afforded words is greater than for non-afforded words. The difference between realistic words and non-afforded words was significant [$t(14) = 2.137$, $p < 0.05$], and the difference between afforded and non-afforded was moderately significant [$t(14) = 1.8$, $p = 0.08$]. The difference between realistic and afforded words was not

significant [$t(14)=0.54, p>0.1$], but the trend is in the right direction. When the raw co-occurrence representation is used, however, the pattern changes: the average cosine for the non-afforded words was statistically equal to afforded words [$t(14) = 0.064, n.s.$]. In addition, unlike the constructed perceptual representations, realistic and non-afforded words did not differ [$t(14) = 1.56, p > 0.1$].

Simulation 2.2: Sensory/motor based priming

Similar to the previous experiment, Myung, Blumstein, & Sedivy (2006) tested whether facilitation occurred when a target word was primed by a word that has sensory/motor based functional information in common with the target, but not associative information (e.g. ‘typewriter’ preceded by ‘piano’). The prime-target pairs focused on manipulation knowledge of objects (e.g. what one can do with a given object). Using a lexical decision task, Myung, et al. found significant facilitation in this condition.

To simulate their experiment, we used the same prime-target word pairs from Myung, et al. (2006) and the same unrelated primes. Because some of the words in this experiment were compounds (‘baby carriage’, ‘safety pin’, etc...), they were transformed to single words (‘carriage’, ‘pin’). Where this changed the meaning of the concept, the word pair was removed from the test. This procedure resulted in 23 word pairs being tested, with each pair having both a related-target and unrelated-target condition. Priming was computed in the model as the related-target cosine minus the unrelated-target cosine.

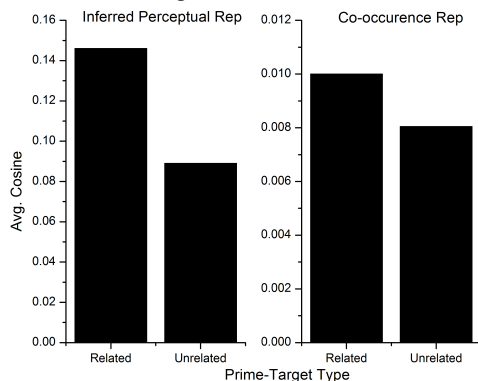


Figure 4. Simulation of perceptual priming results.

The magnitude of priming was assessed for both the inferred perceptual representations and the raw co-occurrence representations. The result of this simulation is depicted in Figure 4, which shows that both representation types show a priming effect. The magnitude of facilitation (*related* > *unrelated*) for the co-occurrence representations was not as pronounced as the inferred perceptual representations, and was not significant [$t(22) = 1.35, n.s.$]. However, the facilitation effect for the inferred perceptual representations was significant [$t(22) = 2.05, p < 0.05$]. This again demonstrates that the perceptual representations inferred by this model contain a considerable amount of knowledge about the perceptual underpinning of words.

Simulation 2.3: Phrase/referent similarity

Wu & Barsalou (2009) had subjects rate the familiarity of novel and familiar noun phrases consisting of a concrete noun preceded by a modifier (e.g. “smashed tomato” vs. “sliced tomato”). Wu & Barsalou argue from the results that conceptual combinations seem to be based on a perceptual simulation of the combined concept. This model is not capable of this advanced simulation process, but we simply wanted to test whether the inferred perceptual representations are better able to account for the familiarity ratings from Wu and Barsalou’s study. Assessing familiarity is the first step to being able to determine conceptual combination, by determining the overlap between the two words’ representations.

The ten novel and ten familiar noun phrases were taken from Wu & Barsalou (2009). Five of the twenty modifiers had to be replaced with their closest synonym (as defined by WordNet) as they were not in the model’s lexicon (due to their very low frequency). To assess familiarity, the cosine between the two words was computed for both the inferred perceptual representation and the raw co-occurrence representation. In addition to examining overall magnitude differences between the conditions, a correlation analysis was conducted over the specific familiarity ratings given to the different noun phrases. Wu & Barsalou published two sets of familiarity ratings: 1) phrase familiarity: how often subject’s had experienced that specific phrase, and 2) referent familiarity: how often subject’s had seen that specific object.

A marginally significant difference was found between the novel and familiar conditions for both the inferred perceptual representations [$t(9) = 2.0, p = 0.07$] and the raw co-occurrence representations [$t(9) = 1.79, p = 0.1$]. However, the item-level fits between the model’s predictions and subject’s familiarity ratings for phrases were also tested. A significant correlation was found between the inferred perceptual representations and subject ratings, for both phrase familiarity [$r = 0.48, p < 0.05$] and referent familiarity [$r = 0.49, p < 0.05$]. However, this was not the case for the co-occurrence representations, as a non-significant correlation was found for both phrase familiarity [$r = 0.12, n.s.$] and referent familiarity [$r = 0.16, n.s.$]. This demonstrates that the inferred perceptual structure is able to simulate item-level variance in familiarity, while the co-occurrence representations are not.

Simulation 2.4: Inferred Modality Representation

As a final simulation, we tested the ability of the model to infer the modality rating data from the Lynott & Connell (2009) norms. In these norms, subjects rate the prominence of the five modalities in representing a target word. As with the McRae et al. (2005) feature vectors, each word was represented as a probability distribution across the five modalities. In Lynott & Connell’s norms, subjects tended to rate vision as consistently more important than other modalities. To reduce this bias in the model, a preprocessing normalization procedure was conducted. Before normalizing

each word vector to a probability distribution, each column was normalized to have a total magnitude of one, which has the effect of standardizing the amount of information that each modality provides. Each word vector was then normalized to a probability distribution.

The model's ability to generate inferred modality ratings was evaluated over a large number of target words from various sources. For the visual, auditory, and tactile modalities the words were taken from van Dantzig, et al. (2008), who conducted a property verification study on these modalities. For the olfactory modality, words were taken from Gonzalez, et al. (2005) who found an increase in activation in olfactory brain regions to words that have a strong smell association. Gustatory words were taken from Goldberg, et al. (2006) who found greater activation in the orbitofrontal cortex to food words. In order to model this, the strength of the proposed modality was measured for each word, and compared against a comparison set of words drawn randomly from another modality. The results of this simulation are displayed in Figure 5. All differences among groups are significant. This demonstrates that this model is able to create correct inferences about the modality basis of words, given a limited amount of starting information.

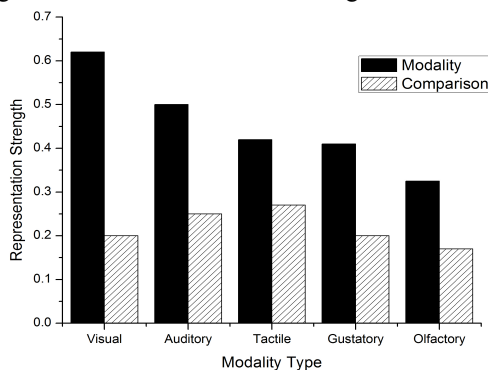


Figure 5. Level of strength for different modalities.

Discussion

We proposed a simulation process, similar in spirit to that suggested by physical symbol systems, to generate inferred perceptual representations for words through the use of global lexical similarity. The perceptual representations are constructed by integrating the already formed (either learnt or inferred) representations of other words, and these are weighted by the scaled associative strength among words in the lexicon. Across many words this simulation process produces a stable representation containing useful perceptual information about how the referent of the word is used. The model was capable of using multiple norm sets, which in turn allowed for a diverse set of data to be tested. The power of this model is not in complex inference or learning mechanisms, but instead is contained in the structure of lexical memory, which has been shown to be an important information source in cognitive modeling (Johns & Jones, 2010).

This model is obviously in the very early stages as an attempt to integrate PSS and distributional models of lexical

semantics. As such, there are currently many shortcomings. One major issue is that the only “perceptual” features that may be inferred are fixed to those used to describe the 541 concrete nouns normed by McRae et al. (2005), which may make it difficult to generalize those features to other types of words in the lexicon. While this shortcoming is no different than other attempts to integrate perceptual and linguistic information (e.g. Andrews et al., 2009), it is rather inflexible (and clearly wrong) to believe that the ~2,500 features generated by McRae et al.’s subjects are sufficient to describe the perceptual structure of the entire lexicon. In addition, the model is subject to making errors of “illusory feature migrations” (Jones & Recchia, 2010); e.g., inferring that *honey* has wings. Nonetheless, the phrasal priming simulations demonstrate that this type of information migration affords the model sufficient power to simulate difficult effects in grounded cognition. Furthermore, the model takes important steps towards the integration of distributional models, global memory frameworks, and creates links to theories of grounded cognition.

References

- Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, 116, 463-498.
- Barsalou, L. W. (1999). Perceptual symbol systems. *BBS*, 22, 577-660.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.
- Glenberg, A. M. & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, 43, 379-401.
- Gonzalez, J., et al. (2006). Reading cinnamon activates olfactory brain regions. *NeuroImage*, 32, 906-912.
- Goldberg, R. F., Perfetti, C. A., Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Neuroscience*, 26, 4917-4921.
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Johns, B. T., & Jones, M. N. (2010) Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, 17, 662-672.
- Jones, M. N., & Riordan, B. (2010). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, 3, 303-345.
- Jones, M. N., & Recchia, G. (2010). You can't wear a coat rack: A binding framework to avoid illusory feature migrations in perceptually grounded semantic models. *Proceedings of the 32nd Annual Cognitive Science Society*.
- 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*, 211-240.
- Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41, 558-564.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, 37, 547-559.
- Myung, J., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: manipulation knowledge of objects. *Cognition*, 98, 223-243.
- Perfetti, C. (1998). The limits of co-occurrence: Tools and theories in language research. *Discourse Processes*, 25, 363-377.
- Recchia, G. L., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information to latent semantic analysis. *Behavior Research Methods*, 41, 657-663.
- Riordan, B., & Jones, M. N. (2010). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*.
- van Dantzig, S., Pecher, D., Zeelenberg, R., & Barsalou, R. W. (2008). Perceptual processing affects conceptual processing. *Cognitive Science*, 32, 579-590.
- Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, 40, 183-190.
- Wu, L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, 132, 173-189.