Representing Abstract Words and Emotional Connotation in a High-dimensional Memory Space

Curt Burgess
Department of Psychology
University of California, Riverside
Riverside, CA 92521
curt@doumi.ucr.edu

Kevin Lund
Department of Psychology
University of California, Riverside
Riverside, CA 92521
kevin@locutus.ucr.edu

Abstract

A challenging problem in the computational modeling of meaning is representing abstract words and emotional connotations. Three simulations are presented that demonstrate that the Hyperspace Analogue to Language (HAL) model of memory encodes the meaning of abstract and emotional words in a cognitively plausible fashion. In this paper, HAL's representations are used to predict human judgements from word meaning norms for concreteness, pleasantness, and imageability. The results of a single-word priming experiment that utilized emotional and abstract words was replicated. These results suggest that it is unnecessary to posit separate lexicons to account for dissociations in priming results. HAL uses global co-occurrence information from a large corpus of text to develop word meaning representations. Representations of words that are abstract or emotional are formed no differently than concrete words.

Any theory of meaning is faced with the problem of what approach to use to represent the meaning of a word. This problem is perhaps even more explicit with computational theories of meaning representation. One approach to modeling word meaning is simply to have binary word vectors where items that are more related are more similar in their pattern of vector elements. Masson (1995) and Plaut (1995) have used this approach. Vector elements in a word have no particular meaning, nor are particular types of word meanings represented. The advantage of this approach is that it is straightforward to model effects that rely on overall patterns of similarity. Other models have used sets of semantic microfeatures intended to instantiate meaning by using a vector of meaningful elements that can collectively be used to derive the meaning of a word. For example, McClelland and Kawamoto (1986) used a set of features in which the concept boy is represented by the features *human, soft, male, medium size, 3-d form, pointed, unbreakable,* and *animate*. Both of these approaches have several drawbacks. It is difficult to represent a large number of concepts. It is also difficult to articulate a set of semantic features, and there is no clearly specified way to represent abstract concepts.

A notably difficult problem is how to represent abstract concepts in computational semantics. The problem has had very little attention probably due to the seemingly intractable nature of developing the representations. In this paper, we argue that language statistics (i.e., global co-occurrence, see Burgess & Lund, 1997) can make a credible first pass at this representational problem. Our model of representing meaning in memory, the Hyperspace Analogue to Language (HAL) model, has been used to account for a range of semantic and associative priming effects (Lund & Burgess, 1996, Lund, Burgess, & Atchley, 1995; Lund, Burgess, & Audet, 1996), the type of semantic errors made by patients with deep dyslexia (Buchanan, Burgess, Lund, 1996), grammatical class distinctions and semantic effects on syntactic processing (Burgess & Lund, 1997) and cerebral asymmetries in semantic memory processing (Burgess & Lund, in press).

HAL is a model that develops word meaning from global co-occurrence statistics by moving a ten-word window along a ~300 million word corpus of Usenet text. Each time the window moves one step (one word), weighted co-occurrence information is tabulated in a large matrix of co-occurrences. Words in the window are recorded as co-occurring with a strength inversely proportional to the number of other words separating them within the window. Words close together receive higher co-occurrence values than words further apart. The process of moving the window and recording these co-occurrences allows for the formation of a co-occurrence matrix. This matrix is 70,000 square since there are 70,000 vocabulary items for which information was stored. Describing the complete details of the model is beyond the scope of this paper and has been done elsewhere (Lund & Burgess, 1996; see Burgess & Lund, 1997, for an update).

The meaning acquisition procedure just described is the same for any word. In this paper, we explore the possibility that HAL's word representations can be used to capture the meaning of abstract words, just as they have been demonstrated to do for more concrete words. We also test whether or not HAL's representations can be used to account for priming effects with words with emotional connotations.

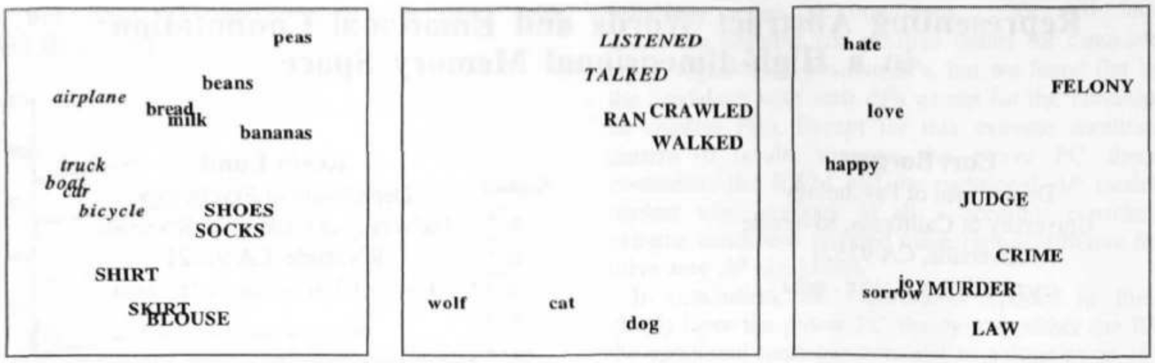

Figure 1: (a) concrete nominal concepts, (b) grammatical concepts, and (c) abstract concepts

Simulation 1: Demonstrating Semantic and Grammatical Categorization

Semantic and grammatical categorization has been demonstrated before using HAL (Burgess & Lund, 1997). The goal of this simulation was to replicate those results with different stimuli and to provide new evidence that abstract concepts can be categorized. In this first simulation, various categories of words are subjected to multidimensional scaling (MDS) in order to show that the interword distances in the high-dimensional space can provide a basis for categorization.

Methods

Co-occurrence vectors were extracted for a set of words, and three MDS solutions were computed. In the first MDS, three categories of nouns were selected (foods, vehicles, & clothing items) to demonstrate, that among concrete concepts, the word vectors provide categorical information sufficient to distinguish among these groups. Vectors were extracted for these words, and, treating each vector as a set of coordinates in a high-dimensional Euclidean space, a distance matrix was formed. Our hypothesis was that this distance matrix, representing the interword distances for the chosen set of words, would operate as a similarity matrix. Each element in the similarity matrix represented the distance between two of the chosen words in this high-dimensional space. The second MDS shows that the word vectors are sensitive to the noun-verb distinction. The third MDS is used with abstract concepts (emotional terms & legal terms).

Results

These three matrices were analyzed by a MDS algorithm that projects points from a high-dimensional space into a lower-dimensional space in a non-linear fashion that attempts to preserve the distances between points (see Manly, 1986, for a discussion of MDS). The lower-dimensional projection allows for the visualization of the spatial relationships between the co-occurrence vectors. These two-dimensional MDS solutions are shown in Figure 1.

Concrete nominal concepts. Visual inspection suggests that the three categories of nouns were differentiated by the MDS procedure (see Figure 1a). Foods, clothes, and vehicles appear in separate spaces. However, given the extreme dimensionality reduction that takes place when doing a two dimensional MDS from high-dimensional data, it was important to conduct inferential statistics. An analysis of variance was computed comparing intragroup distances to intergroup distances. In order to do this, distances between all combinations of item pairs within a group were calculated and compared to all combinations of the group items and all other items. Clothes were differentiated from the other two groups of items, $F(1, 68) = 26.33$, $p < .0001$. Likewise, food concepts were different from clothes and vehicles, $F(1, 68) = 19.17$, $p < .0001$, and the intragroup distances for vehicles differed from the other two groups, $F(1, 68) = 76.98$, $p < .0001$. The two dimensional MDS solutions of the 140,000 dimensional space would seem to be a visually reasonable reduction of dimensionality.

Grammatical concepts. Visual inspection suggests that these three categories of words were also differentiated by the MDS procedure (see Figure 1b). Verbs seem separate from the nouns. Likewise, the communicative verbs (*listened, talked*) differentiate from the movement verbs (*ran, crawled, walked*). An analysis of variance was computed using the high-dimensional vectors comparing intragroup distances to intergroup distances. In order to do this, distances between all combinations of item pairs within a group were calculated and compared to all combinations of the group items and all other items. Nouns (the animals) were differentiated from the verbs, $F(1, 19) = 6.22$, $p < .022$. The communicative verbs were in a different space from the nouns and the movement verbs, F (1, 12) = 5.86, p < .032, and the intragroup distances for the movement verbs differed from the other two groups, F (1, 19) = 6.12, p < .023.

Abstract concepts. As with the nominal categories and the grammatical categories, visual inspection of the two abstract categories suggests that these categories were also differentiated by the MDS procedure (see Figure 1c). The words in the emotional category (e.g., *love, hate, happy*) separate from the words in the legal category (e.g., *judge, law, fair*). An analysis of variance was computed using the high-dimensional vectors comparing intragroup distances to intergroup distances. As with the previous analyses, distances between all combinations of item pairs within a group were calculated and compared to all combinations of the group items and all other items. The emotional words were differentiated from the legal terms, F (1, 43) = 7.00, p < .011. Likewise, the legal terms were different from the emotional words, F (1, 43) = 11.97, p < .0012.

Discussion

The results from these simulations demonstrate that words with similar meanings tend to be close to each other in the high-dimensional space. Vector representations also carry important grammatical-class information which is sufficient to allow one to distinguish between grammatical entities. Although the MDS showed that nouns were distinguished from verbs, internal semantics were maintained as well. For example, *listened* and *talked* clustered together, and the movement verbs clustered together as well. HAL's vector representations were also able to distinguish between two categories of abstract concepts. It has always been problematic with semantic feature systems to articulate a set of features for abstract concepts. Using a global co-occurrence procedure such as HAL's that maps the representation onto the system's experience with the incoming input avoids this problem (see Burgess & Lund, 1997, for more discussion on this issue). As a result, abstract concepts are represented using the same acquisition procedure as concrete concepts.

Simulation 2: Predicting Human Ratings

The MDS results suggest that HAL's representations are as robust for abstract words as they have proven to be for our earlier studies using more concrete words (Lund et al., 1995, 1996) in simulations of semantic priming experiments. A more compelling demonstration, however, would require a larger number of stimuli than shown in Figure 1. In order to accomplish this we used the word vectors to predict several abstract variables made by human raters for words compiled in Toglia and Battig's (1978) *Handbook of Semantic Word Norms.*

Methods

Full-length co-occurrence vectors (140,000 elements) were extracted for 1,486 of ~2,500 words from Toglia and Battig (1978). Very low frequency items were not included in this analysis, since, in order for a word to be in HAL's lexicon, it had to occur at least 50 times in the ~300 million word corpus. This formed a matrix with 1,486 rows and 140,000 columns; column variances were computed, and the thousand most variant columns retained. The resulting matrix was subjected to a principal components analysis in order to concentrate a maximum amount of variance into a minimum number of columns (yielding a new set of vector elements for each word). The most informative 140 components were retained. This procedure yielded, for each word to be analyzed, 140 vector elements. Toglia and Battig had human raters make judgements on each word using a seven point likert scale for three abstract variables (concreteness, pleasantness, and imageability). The word vector elements were used as predictors for the human ratings compiled by Toglia and Battig (1978). Regressions were separately conducted for the three abstract ratings. The score for the ratings for these two variables was predicted by the 140 vector elements (140 predictor elements were used in order to avoid having a ratio of cases to predictors below ten).

Results and Discussion

Reliable predictions were obtained for each abstract variable (concreteness, pleasantness, and imageability). A substantial amount of variance in the human ratings was accounted for by the word vectors: concreteness (R^2 = .65, p < .0001), imageability (R^2 = .63, p < .0001), and pleasantness (R^2 = .28, p < .0001). These results show that the global co-occurrence information carried in the word vectors can be used to predict a tangible proportion of the human likert scale ratings collected for a large set of items on three abstract semantic dimensions. This is an important set of results because they demonstrate the cognitive plausibility of the vector contents (given that they predict human ratings of various dimensions of abstract semantics) moreso than a simple MDS presentation.

Simulation 3: Emotional Semantic Priming

Representing abstract concepts is typically a difficult endeavor for a model of meaning. Likewise, representing emotional connotation is difficult to conceptualize. Dyer (1987) presented three AI approaches to modeling emotion utilizing a highly symbolic, rule-based system that analyzed goals and beliefs of an actor in a situation in order to infer emotional reactions. The drawback to this approach is that an "emotion" expert system is difficult to put forth as a representational model. Perhaps emotional words are just a representational variant of abstract words. This issue was investigated by Bauer and Altarriba (1996) in two semantic priming experiments manipulating the conditions under which abstract and emotional words are presented. Since the regression analyses in Simulation 2 suggest that HAL's representations carry substantial information that can be used to predict abstract dimensions, we decided to investigate the plausibility of accounting for the set of emotional priming results reported by Bauer and Altarriba using the HAL representational model.

Bauer and Altarriba (1996) used a single word priming procedure to investigate four different types of word relationships. They combined as prime-target pairs abstract words that varied as a function of whether the word exceeded a normative level of emotionality. For example, *freedom-liberty* is an abstract-abstract pair (AA); *excited-anxious* is an emotion-emotion pair (EE), in addition to being abstract. They also included incongruent conditions (AE) and (EA). It is important to note that these incongruent pairs were still rated as related as the congruent pairs (incongruent AE example, *fantasy-ecstasy*). Each of these four conditions also had its own set of unrelated controls that were formed by re-pairing the related list such that the prime-target pairs were no longer similar in meaning. The emotion/abstract relationship was maintained however. Thus, for example, priming for the AA condition would be calculated by subtracting lexical decision RTs for this condition from the unrelated AA (AAU, e.g., *clearly-liberty*) condition. Bauer and Altarriba hypothesize that abstract and emotional words reside in separate lexicons. Evidence for this, they argue, would be to find a dissociation in priming for these conditions - a logic that was followed earlier for hypothesizing different memory stores for abstract and concrete words (Bleasdale, 1987).

They conducted two single-word priming experiments. A trial consisted of a fixation point, a prime, followed by a target to which the subject made a lexical decision. In both experiments the target remained on until the subject responded. In the first experiment, the unmasked prime was presented for 200 msec; in the second experiment, the prime was presented for 125 msec and masked for 168 msec. In both experiments, mean lexical decision RTs for the four word relationships (AA, EE, AE, EA) were compared to their unrelated control conditions to calculate the magnitude of semantic priming. In the experiment with the unmasked prime, all four conditions resulted in robust priming effects (ranging from 20 to 26 msec). When the prime was masked, only the AA and EE conditions produced reliable priming effects. They argue that the masked prime condition eliminated the post-access strategic processing and that only when abstract or emotional words are congruently paired can automatic, bottom-up facilitation occur. Furthermore, they conclude that since the still related, but emotion/abstract incongruous conditions did not result in facilitation with masked primes, the abstract and emotional word information is stored in separate lexicons. Automatic priming occurs for items of the same semantic type, but not across types.

We have presented elsewhere a set of results demonstrating that the distances in HAL's high-dimensional memory space reflect the initial bottom-up availability of conceptual information. A range of experiments have been conducted, all converging on the conclusion that HAL's distance metric provides a robust characterization of a variety of semantic priming effects (Buchanan et al., 1996; Burgess & Lund, in press; Lund & Burgess, 1996; Lund et al., 1995, 1996). With HAL, we argue that information related to meaning can be represented in one store.

Methods

Context distances were calculated using the HAL model for all prime-target pairs used by Bauer and Altarriba (1996) in order to obtain a measure of semantic relatedness. These items comprised four related conditions (congruent: AA, EE; incongruent: AE, EA) and their associated unrelated control conditions. There were 20 prime-target pairs for each condition. The unrelated control condition was formed by re-pairing the related items (these were the same unrelated trials used by Bauer and Altarriba). Our methodology assumes that human response latencies to targets in a single-word priming paradigm have some correspondence to the distances between primes and targets in HAL. We have shown elsewhere that HAL does provides robust estimates of human RTs in this task (Lund & Burgess, 1996; Lund et al., 1995, 1996). Thus, an analysis of distances in the high-dimensional meaning space allows us to make an analogous comparison using HAL to how Bauer and Altarriba discussed their priming results.

Results and Discussion

Planned comparisons involve assessing for a priming effect with the context distance metric in the four relatedness conditions (AA, EE, AE, EA) by comparing each with their respective unrelated controls. The pattern of results in the simulation analysis is strikingly similar to the pattern of results Bauer and Altarriba found in their masked prime experiment. (Note: context distances in HAL are scaled to appear similar to RTs in a priming task, see Lund & Burgess, 1996; all distances are in Riverside Context Units or RCUs). Priming was obtained in the AA condition (AAU AA, 647 547 = 100), $F(1,39) = 4.76$, $p < .035$. Priming was also obtained in the EE condition (EEU EE, 653 545 = 108), $F(1,39) = 9.98$, $p < .0032$. Priming was not found in either incongruent condition: AE condition (AEU AE, 663 617 = 46), $F(1,39) = 2.02$, $p < .164$; EA condition (EAU - EA, 669 - 661 = 8), $F(1,37) = 0.09$, $p > .75$.

The simulation results directly parallel those of Bauer and Altarriba. Both congruent conditions produced priming effects; both incongruent conditions did not. It is important to keep in mind that although the incongruent conditions were related (e.g., *fantasy-ecstasy*), they did not produce priming in either their experiment or our simulation. Since these related concepts did not produce priming (in their masked prime, automatic priming experiment), Bauer and Altarriba suggest that this dissociation is due to the priming process not being able to cross these separate lexicons. HAL is a representational model that encodes all words in a similar fashion and represents all words in one memory system, specifically a set of vectors that represent words in the high-dimensional meaning space. Since the simulation with HAL replicates their results, we argue that the separate lexicon hypothesis is unparsimoneous and that Bauer and Altarriba do not get priming across word types simply because the incongruent pairs are not as related as those in the congruent conditions.

General Discussion

The results from the first simulation demonstrate that words with similar meanings tend to be close to each other in HAL's high-dimensional space. These results replicate our earlier semantic and grammatical categorization results (Burgess & Lund, 1997). These MDS results are very similar to the results obtained by Osgood (Osgood et al., 1957) who used a semantic differential procedure. There is a long history in cognitive psychology of using rating scales with some predetermined set of adjectives as bipolar anchors that essentially determine a set of "semantic features" (Smith et al., 1974). HAL does not use a set of "semantic" features as do most models; vector elements correspond to symbols that actually occurred in the input stream. The numeric value of the vector element is a function of the weighted co-occurrence.

Whatever the vector elements correspond to in a model, a commitment is made in using those vectors elements to some view of how environmental information can mapped into a vector. HAL is no exception; one point on which HAL differs from many other models is the range of cognitive effects that have been modeled by using a single type of word representation. Using global co-occurrence information for categorization purposes also avoids the need to recruit raters to make decisions on the array of rating scales that would be required for each word. More recently, Finch and Chater (1992) demonstrated that words can be categorized according to grammatical class using a learning procedure very similar to ours. The results from this simulation suggest that the vector representations carry important grammatical-class information which is sufficient to allow one to distinguish between grammatical entities. Although the MDS showed that nouns were distinguished from verbs, internal semantics were maintained as well. For example, *listened* and *talked* clustered together, and the movement verbs clustered together.

HAL's vector representations were also able to distinguish between two categories of abstract concepts. It has always been problematic with semantic feature systems to articulate a set of features for abstract concepts. Using a global co-occurrence procedure such as HAL's that maps the representation onto the system's experience with the incoming input avoids this problem. As a result, abstract concepts are represented using the same acquisition procedure as concrete concepts. The results from these simulations are consistent with our earlier idea (Lund et al., 1995) that the categorical information carried in the vector representations emerges, in some abstract way, due to the substitutability of words in contexts. Thus, grammatical-categorical knowledge and semantic-categorical knowledge would seem to be an integral part of this substitutability process, as well as being implicit in HAL's vector representations. The vector representations appear to encode information that provides basic categorical (semantic and grammatical) knowledge. In addition, it would appear that the meaning acquisition process allows for the encoding of abstract as well as concrete knowledge.

In Simulation 2, the word vectors were used to predict several abstract variables made by human raters for words compiled by Toglia and Battig (1978) in the *Handbook of Semantic Word Norms*. These results show that the global co-occurrence information carried in the word vectors can be used to predict human likert scale ratings collected for a large set of items on three abstract semantic dimensions. The results from this simulation are important because we argue that the vector representations are cognitively plausible. The regressions demonstrated that the vector contents are able to account for a substantial amount of the variance in predicting the human ratings of the three abstract semantic dimensions.

The results from the third simulation that modeled the emotional word priming study of Bauer and Altarriba (1996) were particularly compelling. Both of the congruent conditions produced priming effects; both incongruent conditions did not. This pattern of results was obtained by Bauer and Altarriba and in our simulation. Bauer and Altarriba argue that this dissociation is due to the priming process not being able to cross these separate lexicons. We think that the separate lexicon hypothesis is unparsimoneous given our results, however, since HAL is a representational model that encodes all words using the same acquisition procedure and represents all words in one memory system. The results of Simulation 3 lead to the conclusion that Bauer and Altarriba do not get priming across word types because the words in the incongruous condition are not as related as the congruent pairs.

The series of results presented in this paper extend the range of effects that the HAL model has been able to mimic. We think HAL is best viewed as a model of subconceptual meaning the meaning that is initially accessed during word recognition. This is why HAL provides an account of such an extensive range of priming results. HAL's limitations, and the reason that we do not claim that it is a complete model of all meaning, are concerns with higher-level semantic computation. However, a model of meaning should be able to represent emotional connotation. Emotional words are used frequently in language and play an important role in the communicative process. We certainly are not claiming that HAL's representations completely characterize emotional connotation. However, the representations do seem to capture sufficient emotional meaning in abstract words such that several set of data from experiments with human subjects were able to be simulated with HAL. These results and the ability to represent abstract words and emotional connotation with the same meaning acquisition procedure as used for any other word go a long way towards being able to make a claim for HAL being a more general model of meaning representation.

Acknowledgments

This research was supported by NSF Presidential Faculty Fellow award SBR-9453406 to Curt Burgess. We thank Catherine Decker and Christine Chiarello for their helpful comments. We especially thank Michael Toglia and William Battig for supplying a machine-readable version of their handbook used in Simulation 2, and Lisa Bauer and Jeanette Altarriba for supplying us with their stimuli used in Simulation 3 and for their helpful comments.

References

Bauer, L. M., & Altarriba, J. (1996). *Emotion words are not abstract.* Paper presented at the Psychonomic Meeting, Chicago, IL.

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 582-594.

Buchanan, L., Burgess, C., & Lund, K. (1996). Overcrowding in semantic neighborhoods: Modeling deep dyslexia. *Brain and Cognition, 32*, 111-114.

Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimensional context space. *Language and Cognitive Processes, 12*, 1-34.

Burgess, C., & Lund, K. (in press). Modeling cerebral asymmetries of semantic memory using high-dimensional semantic space. In Beeman, M., & Chiarello, C. (Eds.), *Right hemisphere language comprehension: Perspectives from cognitive neuroscience.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Dyer, M. G. (1987). Emotions and their computations: Three computer models. *Cognition and Emotion, 1*, 323-347.

Finch, S., & Chater, N. (1992). Bootstrapping syntactic categories by unsupervised learning. *Proceedings of the Cognitive Science Society* (pp. 820-825). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lund, K., Burgess, C., & Audet, C. (1996). Dissociating semantic and associative word relationships using high-dimensional semantic space. *Proceedings of the Cognitive Science Society* (pp. 603-608). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Lund, K., Burgess, C., & Atchley, R.A. (1995). Semantic and associative priming in high-dimensional semantic space. *Proceedings of the Cognitive Science Society* (pp. 660-665). Hillsdale, NJ: Lawrence Erlbaum Associates.

Manly, B. F. J. (1986). *Multivariate statistical methods: A primer.* London: Chapman & Hall.

Masson, M.E.J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 3-23.

McClelland, J.L., & Kawamoto, A.H. (1986). Mechanisms of sentence processing: Assigning roles to constituents. In Rumelhart, McClelland, and the PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition: Volume 2: Psychological and biological models.* Cambridge: MIT Press.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). The measurement of meaning. Urbana, IL: University of Illinois Press.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society* (pp. 37.42). Hillsdale, NJ: Lawrence Erlbaum.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Lawrence Erlbaum Associates.