Semantic Similarity Priming Without Hierarchical Category Structure

Ken McRae (kenm@sunrae.sscl.uwo.ca)
George S. Cree (george@sunrae.sscl.uwo.ca)
Chris McNorgan (cmn@rogers.wave.ca)
Department of Psychology
University of Western Ontario
London, Ontario, Canada

Abstract

In an attractor model of semantic memory, semantic similarity is determined by degree of featural overlap. In contrast, in spreading activation theory, two concepts are similar if they share features or if they are linked to the same superordinate category node. We present an attractor network model of computing word meaning and use it to simulate the data of McRae and Boisvert (in press), who found that short SOA semantic similarity priming directly depends on degree of featural overlap. The two accounts of semantic similarity are then contrasted in a human experiment. In support of attractor networks, priming effects were determined by featural overlap, and no evidence was found for priming through a purported superordinate node. It is concluded that lexical concepts are not represented as static nodes in a hierarchical system.

Semantic Similarity

Semantic memory research has been dominated by semantic networks and the associated spreading activation theory (Collins & Loftus, 1975). Recently, however, theories of the computation of word meaning have been expressed in terms of distributed attractor networks (Hinton & Shallice, 1991). Semantic networks are typically hierarchical in nature, with categories at different levels represented by individual nodes. Category membership is thus explicitly coded via links between exemplar and category nodes. In contrast, category membership is not coded in such an explicit manner in attractor networks. The present work focuses on this contrasting aspect of these two theories, testing this difference in the realm of semantic similarity priming.

Semantic similarity priming refers to the fact that response latency to a target word such as *hawk* is faster when it is preceded by a similar word such as *eagle* versus an unrelated word such as *bread*, even when the similar prime and target are not normatively associated. Priming research has played a key role in the development and testing of theories of semantic memory because most researchers believe that results of these experiments directly reflect the structure of semantic memory, particularly when subjects' strategies are minimized.

The mechanisms used to account for semantic similarity priming differ in spreading activation models versus attractor networks. In spreading activation theory, recognizing a word includes activating its corresponding node in semantic memory. Response latency is assumed to be directly related to the time required for a word's node to reach an activation threshold. Critical for explanations of priming, when a word is activated, activation spreads from it to all linked nodes. The existence and strength of these links is assumed to directly reflect learned semantic relationships between pairs of words. Similarity-based priming is due to two mechanisms, the first being spreading activation from prime to target via links between shared features (e.g., *eagle* ⇒ <has wings> ⇒ *hawk).* The second is that activation is assumed to spread via "highly criterial" links connecting exemplar and category nodes (*eagle* ⇒ *bird* ⇒ *hawk*) (Collins and Loftus, 1975, p. 413).

In attractor networks, recognizing a word involves settling into a stable state in a multi-dimensional space. Short SOA semantic priming is thought to be due to residual activation of the prime's meaning, which influences the ease with which the model can move from one attractor state to another (Masson, 1995; Plaut, 1995). In a typical simulation, the prime's word form is input to the network and its meaning is computed. With the network thus in a state representing the prime (*eagle*), the target's word form (*hawk*) is given as input. Facilitation results because the distributed semantic representations of the prime and target overlap, so that some portion of the semantic units for the target begin in their correct state. This is not the case when the prime is not related to the target. Critically important is the fact that category membership in these networks is given no special status. Therefore, priming is not due to category membership per se.

The majority of human empirical studies of semantic similarity priming have operationalized similarity on the basis of shared superordinate category (e.g., Lupker, 1984; Moss, Ostrin, Tyler, and Marslen-Wilson, 1995; Shelton and Martin, 1992). However, these studies have found little or no priming between words defined as similar on this metric (but see Chiarello et al., 1990). On the other hand, McRae and Boisvert (in press) clearly demonstrated that similarity priming depends on the degree of featural overlap between two lexical concepts, with strong featural similarity being required for priming to occur. McRae and Boisvert also noted that the studies in which null effects were obtained used prime-target pairs that were only moderately similar, presumably because the researchers felt that shared category membership was key, following spreading activation theory.

Although it appears that featural similarity is the key variable to explain this set of behavioral phenomena, it is not clear whether the notion of shared category membership is necessary to account for similarity priming. Therefore, the goals of this article are to show that an implemented attractor network simulates McRae and Boisvert's (in press) result that degree of featural similarity determines amount of priming, and to directly contrast featural similarity versus shared superordinate category in a human experiment.

This article is structured as follows: (1) we present an attractor model of computing word meaning; (2) via simulation, we demonstrate that short SOA semantic similarity priming effects are best predicted by similarity in terms of featural overlap; and (3) we present an experiment showing that semantic similarity priming effects are influenced by featural similarity, but not hierarchical category structure (typicality).

Model of Computing Word Meaning

The key elements of the model are: (1) semantic representations were derived from subjects, rather than being experimenter-created, so that degree of similarity is not a free parameter; (2) a word's meaning is an attractor point in semantic state space; (3) the mapping from word form to meaning is arbitrary; and (4) there is no explicit hierarchical structure.

The model's architecture is presented in Figure 1. The network mapped directly from 12 word form units to 1242 semantic features. The semantic feature units looped back to themselves through a layer of 30 hidden units (semantic structure units).

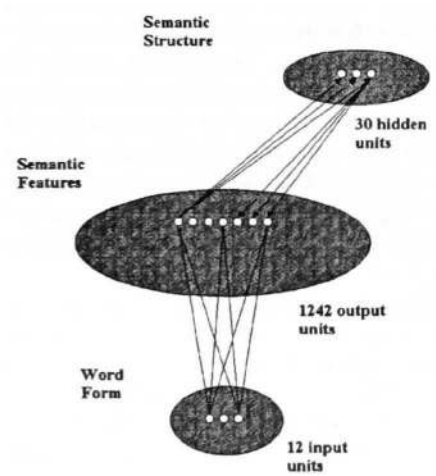

Figure 1: Model Architecture.

Units

Input to the network was an abstract word form representation that could be interpreted as either spelling or sound. Each word's form was represented by turning on a randomly selected 3 units (activation = 1). Thus, the mapping between meaning and form was arbitrary, as in English monomorphemic words.

The semantic representations were taken from McRae, de Sa, and Seidenberg (1997), who asked subjects to produce semantic features for 19 exemplars from each of 10 object categories: birds, mammals, fruits, vegetables, clothing, furniture, kitchen items, tools, vehicles, and weapons. The resulting output representation was sparse because a concept consisted of at most 27 features across the semantic feature layer.

McRae et al. (1997) noted that relevant structure is not restricted to mappings between domains such as orthography, phonology, and semantics, but also includes structure within a domain. Thus, the semantic structure units play an important role as hidden units in that they encode semantic regularities (feature correlations) and exploit these regularities for computing word meaning. In essence, this cyclical part of the network is where the attractors are formed, so that the model's computational dynamics are strongly influenced by correlations among semantic features, such as <has wings> and <has a beak>.

Training

The semantic and semantic structure units were initialized to random starting values in the range 0.2 ± .05. A word's form was then hard-clamped at the input layer. Each tick of processing time (20 in total) allowed activation to spread one layer forward. Total time was segmented into 4 time steps (t), each consisting of 5 time ticks ($\tau = 0.2$) (similar to Plaut, 1995). Net inputs to a unit (x_j) were averaged according to Equation 1,

$$(1) \qquad x_j^{[t]} = \tau \sum_i s_i^{[t-\tau]} w_{ij} + (1-\tau) x_j^{[t-\tau]}$$

where s_i was the activation of the i^{th} unit and w_{ji} was the weight on the connection to the j^{th} unit from the i^{th}. Therefore, a unit's input at each time step was a weighted average of its previous input and the current input from all sending units. Activation was then determined using the standard sigmoidal function.

Weight changes were calculated using the backpropagation-through-time learning algorithm. Error was backpropagated over the 20 processing ticks in a manner analogous to the forward pass. Error was injected into the system (i.e., the target's semantic representation was provided) for the final two time steps only (10 time ticks), thereby training the network to produce the target output gradually over time. Error derivatives were calculated using cross entropy error (E) as in Equation 2,

$$(2) \qquad E = \sum_p \sum_i d_i \ln y_i + (1-d_i)\ln(1-y_i)$$

where d_i was the desired activation for $unit_i$ and y_i was the computed activation, summed over patterns p.

The network was trained using the PDP++ (version 1.1) simulator developed at Carnegie Melon by R. C. O'Reilly, C. K. Dawson, and J. L. McClelland. Weights were updated after each pattern presentation. The learning rate was 0.01 throughout training. Momentum was set at 0 for the first 10 epochs of training and 0.9 thereafter. Each epoch consisted of randomly presenting the 190 patterns. After 85 epochs of training, the network settled to the correct stable state for all patterns within 20 time ticks.

Simulation: McRae and Boisvert (in press, Experiment 3)

The model was used to simulate Experiment 3 of McRae and Boisvert (in press), in which it was demonstrated that semantic similarity priming is crucially dependent on the degree of featural overlap. They designed word triplets by pairing a target (*jar*) with both a highly similar prime (*bottle*) and a less similar prime (*plate*), with degree of similarity being established by subjects' ratings. Prime-target similarity of the less similar primes was in the range of Shelton and Martin's (1992) items. SOAs of 250 ms and 750 ms were used because short SOAs such as 250 ms are believed to reflect lexical-internal factors only, whereas effects at a longer SOA such as 750 ms may be influenced by subjects' strategies (see Neely, 1991 for a review). Consistent with these notions, with a 250 ms SOA, latencies in a semantic decision task ("Does it refer to a concrete object?") were faster for targets preceded by highly similar primes (685 ms) than by less similar (712 ms) or dissimilar primes (711 ms), and no priming obtained for the less similar items. With a 750 ms SOA, semantic decisions were again faster in the highly similar condition (646 ms) than in both the less similar (664 ms) and dissimilar conditions (692 ms). However, reliable priming was found for the less similar items, replicating Shelton and Martin's findings.

The simulation investigated whether the network would exhibit appropriate settling for the same items. That is, the model should show faster settling times for targets preceded by highly similar primes versus either less similar or dissimilar primes. We did not attempt to closely approximate the long SOA condition because it is unclear how to incorporate subjects' strategies into the network. Of interest, however, is the prediction that less similar targets should converge somewhat faster than dissimilar ones because there must be a basis for the priming obtained at the long SOA.

Method

Prior to presenting the prime, all semantic and semantic structure units remained in the state determined by the previous target. The prime's word form was hard-clamped for 15 ticks. The target's word form was then clamped with all other units unchanged. The target was allowed to settle for 20 ticks and cross-entropy at the semantic feature layer was recorded. Priming trials were run using five random orders, with the results averaged across runs.

Results

The settling profiles for the targets are presented in Figure 2A. The difference between the high and less similar groups is more pronounced than between the less similar and dissimilar items, reflecting the human data. It is not clear how best to map such data onto human decision latencies because of the uncertainty involved in determining the extent to which a representation must stabilize before a response can be initiated in a speeded task. Therefore, we calculated the mean number of time ticks required to reach several levels of cross-entropy, and these results are presented in Figure 2b. Settling times are presented for a number of thresholds, ranging from 2.5 to 0.5 (in decrements of 0.25).

A two-way repeated measures ANOVA was conducted with error level (2.5 - 0.5) and prime-target similarity (highly similar vs. less similar vs. dissimilar) as the independent variables, and number of ticks to reach the specified error level (convergence latency) as the dependent variable. Prime target similarity influenced convergence latency, $F(2,52) = 26.75$[1]. With the nine error levels combined, convergence latency for the highly similar items was significantly shorter than for the less similar, $F(1,52) = 30.35$, and dissimilar items, $F(1,52) = 47.90$, but the less similar and dissimilar targets did not differ, $F(1,52) = 1.99$, $p > .1$.

Similarity and error level interacted in that differences among the three conditions decreased at the lower error levels, $F(16,416) = 17.75$. Planned comparisons revealed that at error levels 2.5 and 2.25, convergence latency for highly similar targets was significantly shorter than for less similar targets, which was in turn shorter than for dissimilar targets. These results mirror those of the long SOA presentation condition of McRae and Boisvert's Experiment 3. At the next 5 error levels (2 to 1), highly similar targets converged more quickly than less similar targets, which did not differ significantly from the dissimilar condition. These results mirror their short SOA. At error levels of 0.75 and 0.5, only the highly similar and dissimilar groups differed

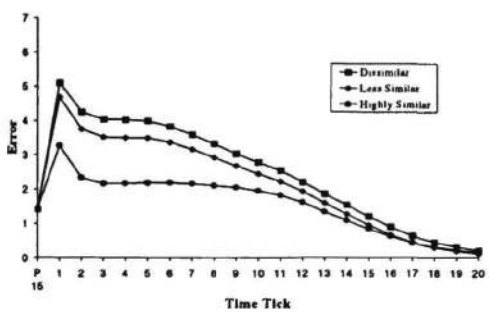

Figure 2a: Mean error at each tick.

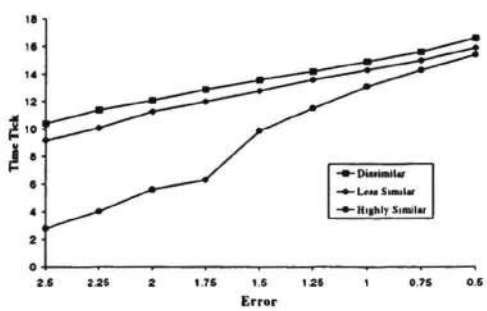

Figure 2b: Mean number of ticks to reach error level.

[1] Note that $p < 0.05$ unless otherwise indicated. Where appropriate, F_1 refers to analyses by subjects whereas F_2 refers to analyses by items.

reliably. Finally, convergence latency increased across the nine error levels, $F(8,208) = 147.14$.

Discussion

The priming effects demonstrated by the model reflected the subtle effects seen in the human data. The small difference between the convergence latencies of the less similar and dissimilar conditions explain the small priming effects found by Lupker (1984), Moss et al. (1995), and Shelton and Martin (1992). Note that the degrees of freedom for accounting for these effects were minimized because degree of similarity was determined by McRae et al.'s (1997) feature production norms, rather than by the experimenters' intuitions, and all items in the human experiment were included in the model.

Experiment

McRae and Boisvert's (in press) Experiment 3 and the simulation thereof suggest that featural similarity is the primary determinant of semantic similarity priming effects. In contrast, according to most semantic network models such as Collins and Loftus (1975), semantic priming is also mediated by superordinate category nodes. In this account, the strength of the exemplar ⇔ superordinate links are directly related to exemplar typicality. Thus priming between category co-ordinates should depend on their typicality. To investigate this, Chiarello and Richards (1992) compared priming effects for typical (*robin-crow*) versus less typical (*duck-crow*) members of a category, with the two groups equated for rated featural similarity. Priming was found for the highly typical primes in a lexical decision task when primes and targets were presented in the left visual field, but not when words were presented to the right. In a pronunciation task, numerically but not significantly larger priming effects were found for the highly typical primes in both visual fields. In summary, these experiments are suggestive, but they do not clarify the role of typicality.

To investigate this issue further, we reanalyzed McRae et al.'s (1997) Experiment 3 short SOA priming data. They measured priming effects for 88 items that ranged both in item similarity and in typicality of the primes and targets. Item by item priming effects were predicted using typicality of the prime, typicality of the target, summed typicality of the prime and target, and similarity in terms of individual and correlated features. Two items were deleted because the typicality ratings for *shed-barn* and *crayon-pencil* were collected with respect to the superordinate *tool*, and the ratings showed that subjects did not consider them part of this category. For the 86 prime-target pairs, similarity in terms of correlated feature pairs was the strongest predictor, $r^2 = 0.16$, $F(1,83) = 15.60$, and similarity in terms of individual features also significantly predicted priming effects, $r^2 = 0.15$, $F(1,83) = 14.10$. In contrast, none of the typicality measures predicted priming: prime typicality, $r^2 = 0.03$, $F(1,83) = 2.37$, $p > 0.1$; target typicality, $r^2 = 0.01$, $F(1,83) = 1.10$, $p > .2$; summed typicality, $r^2 = 0.03$, $F(1,83) = 2.15$, $p > 0.1$. Note that because the variation of the typicality ratings was slightly greater than of the two similarity measures, any differences in predictive ability cannot be attributed to this factor.

In the present experiment, similarity and typicality were compared in a more direct fashion. Targets were paired with more similar/less typical ("Similar") and less similar/more typical ("Typical") primes (e.g., *squash* as the target, *pumpkin* as the Similar prime, *corn* as the Typical prime). Extensive norming was conducted to ensure these conditions were met. If featural similarity is the key predictor of short SOA priming effects, as predicted by an attractor network, then priming should obtain only for the Similar prime-target pairs. If priming occurs through a superordinate node and shared features, as in a spreading activation network, it should be relatively equal for both types of items. In this case, featural similarity would play the stronger role for the Similar items, whereas shared superordinate category node would dominate when the primes are Typical.

Norming

Norming studies produced 18 triplets (from 75 candidate triplets) that included a target, a Similar prime, and a Typical prime (see Appendix).

Word Association Norms To ensure that the prime-target pairs were not normatively associated, the experimenter read aloud either the targets or one of the primes (in three lists, 16 subjects each) from the 75 triplets originally constructed. A word triplet was discarded if more than one subject produced the target as a response to either prime, or vice versa. This left 51 nonassociated triplets.

Category Production Norms Forty-five subjects were shown one word from each of the 51 triplets. The experimenter read each item aloud and the subject indicated the category to which she believed the concept belonged. The most frequent response was designated as the item's dominant category. Eighteen triplets were retained on the basis that the dominant category was identical for each of its members and at least 50% of the subjects produced it for each. The mean percentage of dominant category responses, along with other stimuli characteristics, are presented in Table 1. Subjects produced the dominant category name more frequently for Typical primes than for Similar primes, $F_1(1,42) = 8.59$, $F_2(1,34) = 13.85$.

For a few of the 18 items, a secondary superordinate category name was produced by more than one subject. In terms of predicting priming effects, a concern arises if the secondary category is more restricted than the dominant category because it could be the case that priming is mediated by that closer superordinate node. This situation arose for the *waffle - toast - pancake* triplet only, where *food* was the dominant category and *breakfast food* was the less inclusive secondary category. Subjects produced *breakfast food* 33% of the time to *waffle* and *pancake*, and 7% of the time to *toast*, which was the Typical prime.

Typicality Ratings Seventeen subjects rated the typicality of each member of the 18 triplets. Each item was included with its dominant category. Sentences of the form: "How typical of a VEGETABLE is CORN?" were presented along with a 9 point scale, where 1 corresponded to "not at all

typical" and 9 to "extremely typical". Subjects rated the Typical primes as more typical than the Similar primes, $F_1(1,32) = 45.83$, $F_2(1,34) = 4.95$.

Similarity Ratings Thirty-six subjects rated the similarity of the 18 triplets on a 9 point scale, where 1 corresponded to not at all similar and 9 to extremely similar. Subjects rated the Similar primes as being more similar to the targets than were the Typical primes, $t_1(34) = 4.79$, $t_2(17) = 6.43$.

Table 1: Characteristics of Stimuli.

	Target	Similar Prime	Typical Prime
Sim		6.0(0.3)	4.4(0.3)
Dom	72.3(3.5)	70.1(3.3)	83.0(2.4)
Typ	6.6(0.3)	6.8(0.3)	7.8(0.2)
Lets	6.5(0.4)	5.4(0.3)	5.3(0.3)
Freq	7.7(2.7)	10.2(3.7)	16.3(6.5)

Note: (Standard error in parentheses)
Sim = similarity to target; Dom = dominant category response; Typ = typicality; Lets = length in letters; Freq =word frequency (Kucera and Francis, 1967)

Method

Subjects Sixty-six University of Western Ontario undergraduates participated (22 per list) either for course credit or for cash remuneration. All subjects were native speakers of English, and had normal or corrected-to-normal vision.

Materials Three lists were created so that subjects saw no prime or target twice. For each list, 6 targets were paired with Similar primes, 6 with Typical primes, and 6 with unrelated primes. Unrelated trials were created by re-pairing similar primes with the targets. There were 102 filler trials per list, consisting of 42 unrelated word-word and 60 word-nonword pairs. The relatedness proportion was 0.2, and the nonword ratio was 0.56.

Procedure Subjects were tested individually using PsyScope (Cohen et al., 1993) on a Macintosh LC630 with a 14-inch color Sony Trinitron monitor. They responded by pressing one of two buttons on a CMU button box. The subjects' index finger of their dominant hand was used for a "yes" response. A trial consisted of a fixation point "+" for 250 ms, followed by the prime for 200 ms, a mask (&&&&&&&&&&) for 50 ms, and then the target, which remained on screen until the subject made a lexical decision. The ITI was 1500 ms. Subjects were given 40 practice trials followed by 120 experimental trials.

Design The independent variable was prime type (Similar vs. Typical vs. unrelated). A list factor (or item rotation group) was included. Prime type was within subjects and items. The dependent measures were decision latency and accuracy.

Results

Mean decision latency and error rate for each condition are presented in Table 2. Latencies greater than 3 standard deviations above the grand mean were replaced by the cutoff value (1% of the scores).

Lexical decision latencies differed by prime type, $F_1(2,162) = 4.26$, $F_2(2,30) = 4.75$. Planned comparisons revealed that subjects responded 26 ms faster to the Similar pairs than to the unrelated pairs, $F_1(1,162) = 7.34$, $F_2(1,30) = 8.72$. Furthermore, subjects responded 22 ms faster to the Similar pairs than to the Typical pairs, $F_1(1,162) = 5.29$, $F_2(1,30) = 5.02$. The 4 ms priming effect for the Typical pairs was not reliable, $F_1 < 1$, $F_2 < 1$.

No differences were significant in the error data.

Table 2: Mean Decision Latency in ms and % Errors.

Prime Type	Decision Latency	Errors
Similar	636 (12)	6.2 (1.1)
Typical	658 (12)	7.3 (1.2)
Unrelated	662 (12)	9.7 (1.2)

Note: (Standard error in parentheses)

Discussion

Subjects were faster to respond to targets preceded by Similar primes than those preceded by either Typical or unrelated primes, and there was a small nonsignificant difference between the latter two conditions. Thus, semantic similarity priming is a product of featural overlap, rather than shared superordinate category. These results can be taken as evidence to refute one central aspect of most versions of spreading activation theory; semantic memory does not consist of a set of concept nodes organized in a hierarchical fashion. Note that although the hierarchical nature of semantic memory and the key role played by superordinate nodes are critical components of semantic network theories such as Collins and Loftus (1975), this experiment does not completely discount spreading activation models of semantic memory. Even without this mechanism, similarity-based priming effects can be attributed to featural links between highly similar concept nodes, or direct concept-concept links.

In addition to empirical problems, there are logical problems with an account of semantic memory that emphasizes hierarchical semantic structure coded in terms of local category nodes. For instance, there are inherent difficulties in determining what categories are psychologically real, and hence what category nodes would be implicated to play a role in, for example, semantic priming. Many types of concepts exist for which the relevant superordinates are not obvious, particularly to the average person. These might include verbs such as *run* or *break*, adjectives such as *silent* or *beautiful*, and even concrete nouns such as *fence* or *garage*. This problem became painfully apparent when constructing candidate items for the category production task of the experiment in that it was difficult to create basic-level concepts that we felt would induce consistent superordinate category

responses in the absence of any biasing context. Along similar lines, Barsalou (1987) has argued that superordinate categories should not be viewed as static nodes in a hierarchically-organized semantic system. Rather, on the basis of the variation in typicality ratings across individuals and within individuals over time, as well as results showing that people treat ad hoc categories such as *things on my desk* in much the same way as taxonomically-based categories, he concluded that people's representations of categories are not stable entities, but are computed only when needed and are constantly changing as a result of experience.

Conclusions

Evidence was presented to support an attractor network theory of the computation of word meaning by demonstrating that semantic similarity priming effects are best explained in terms of featural overlap, rather than explicitly encoded shared category membership. This work adds to the growing list of phenomena in word recognition that have been accounted for, or predicted by, attractor networks of lexical processing.

Acknowledgements

This research was funded by NSERC grant OGP0155704 to the first author and an OGS Postgraduate scholarship to the second.

References

Barsalou, L. W. (1987). The instability of graded structure in concepts. In U. Neisser (Ed.), *Concepts and conceptual development: ecological and intellectual factors in categorization* (pp. 101-140). Cambridge, Cambridge University Press.

Chiarello, C., Burgess, C., Richards, L., & Pollock, A. (1990). Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't ... sometimes, some places. *Brain and Language, 38*, 75-104.

Chiarello, C., & Richards, L. (1992). Another look at categorical priming in the cerebral hemispheres. *Neuropsychologia, 30*, 381-392.

Collins, A. M. & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82*, 407 - 428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*, 240-247.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, & Computers, 25*, 257-271.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press.

Hinton, G.E. & Shallice, T. (1991). Lesioning an Attractor Network: Investigations of acquired dyslexia. *Psychological Review, 98*, 74-95.

Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior, 23*, 709 - 733.

McRae, K. & Boisvert, S. (in press). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

McRae, K., de Sa, V. & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126*, 99-130.

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

Moss, H. E., Ostrin, R. K., Tyler, L. K. & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 863 - 883.

Neely, J.H. (1991). Semantic Priming Effects in Visual Word Recognition: A selective view of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic Processes in Reading: Visual word Recognition* (p. 264-336). Hillsdale, NJ: Erlbaum.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society, 17*, 37-42.

Postman, L., & Keppel, G. (1970). *Norms of word associations*. San Diego, CA: Academic Press.

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 1191 - 1210.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A feature model for semantic decisions. *Psychological Review, 81*, 214-241

Appendix 1: Prime-target pairs from the Experiment

More Typical Prime	More Similar Prime	Target	Dominant Category
sparrow	eagle	hawk	bird
robin	parakeet	budgie	bird
vulture	duck	chicken	bird
apple	plum	prune	fruit
peach	coconut	pineapple	fruit
corn	pumpkin	squash	vegetable
carrot	radish	beets	vegetable
cucumber	peas	beans	vegetable
toast	waffle	pancake	food
strudel	cupcake	muffin	food
pepper	nutmeg	cinnamon	spice
ocean	stream	creek	body of water
shirt	bra	camisole	clothing
torch	lamp	chandelier	light source
silk	denim	corduroy	fabric
tuba	flute	clarinet	musical instrument
rake	hoe	shovel	gardening tool
gun	missile	bomb	weapon