

## Language and Cognitive Processes, Volume 6, 1991

### Author Index

Barry, G., 229  
 Berg, T., 265  
 Bever, T., 49  
 Campbell, R., 29  
 Charles, W., 1  
 de Partz, M., 137  
 Emmorey, K., 207  
 Gernsbacher, M., 81  
 Hodgson, J., 169  
 Miller, G., 1  
 Norman, F., 207  
 O'Grady, L., 207

Pickering, M., 229  
 Pillon, A., 137  
 Raison, A., 137  
 Rosen, S., 29  
 Seron, X., 137  
 Shanon, B., 339  
 Solis-Macias, V., 29  
 Tanenhaus, M., 303  
 Townsend, D., 49  
 Trueswell, J., 303  
 Webber, B., 107  
 White, T., 29

### List of Reviewers

The Editors wish to thank the following outside reviewers for the time and effort they have given to the journal in 1991:

P. Ayton	E. Funnell	S. Monsell
J. Beattie	S. Garnsey	J. Myers
T. Bever	P. Gordon	E. Newport
S. Brennan	D. Green	J. Oakhill
T. Briscoe	P. Hagoort	T. Shallice
G. Carlson	T. Harley	M. Shatz
J. Cassell	M. Hickmann	K. Stenning
S. Crain	D. Howard	L. Surian
C. Clifton	S. Lipka	J. Trueswell
S. D'Zmura	M. MacDonald	P. Wasow
F. Ferreira	P. Meara	B. Webber
S. Franklin	D. Mitchell	P. Zwitserlood
L. French	M. Moens	

### Contextual Correlates of Semantic Similarity

George A. Miller

Department of Psychology, Princeton University, Princeton, New Jersey,  
U.S.A.

Walter G. Charles

Department of Psychology, Oregon State University, Corvallis, Oregon,  
U.S.A.

The relationship between semantic and contextual similarity is investigated for pairs of nouns that vary from high to low semantic similarity. Semantic similarity is estimated by subjective ratings; contextual similarity is estimated by the method of sorting sentential contexts. The results show an inverse linear relationship between similarity of meaning and the discriminability of contexts. This relation is obtained for two separate corpora of sentence contexts. It is concluded that, on average, for words in the same language drawn from the same syntactic and semantic categories, the more often two words can be substituted into the same contexts the more similar in meaning they are judged to be.

### INTRODUCTION

Although synonymy is a familiar semantic relation between words, defining synonymy is more complicated than it might at first seem. In everyday discourse, two words are said to be synonyms if they have the same meaning, but "the same meaning" is what needs to be defined. Because it is difficult to assess the meaning of a word outside of the context in which it is used, most semanticists prefer to phrase their definition of synonymy in terms of *statements* that have the same meaning, where having

---

Requests for reprints should be addressed to George A. Miller, Department of Psychology, Princeton University, Princeton, New Jersey 08544, U.S.A.

The preparation of this paper was supported in part by Contract MDA 903-86-K-0242 with the Army Research Institute, and in part by a grant from the James S. McDonnell Foundation. The views and conclusions contained herein are those of the authors and should not be represented as official policies of ARI or the James S. McDonnell Foundation. The data were collected by W.G.C. while he was Assistant Professor at the State University of New York at Oswego. The authors are indebted to P. N. Johnson-Laird for his comments on an earlier draft of the manuscript.

## 2 MILLER AND CHARLES

"the same meaning" can be defined as having the same truth values. Following a formulation usually attributed to Leibniz, two words are said to be synonyms if one can be used in a statement in place of the other without changing the meaning of the statement (the conditions under which the statement would be true or false).

But even this more careful definition is subject to misunderstanding. For example, it is somewhat anomalous that synonymy is recognised as a basic semantic relation by lexicographers and linguistic semanticists (Cruse, 1986; Lyons, 1977) at the same time that many philosophical semanticists question whether there are any true synonyms in natural languages (Quine, 1953). Of course, these two positions are not as contradictory as they sound. The philosophers suspect that for any pair of synonyms there are statements whose meanings (truth values) are changed by using one of the words in place of the other. Few lexicographers would disagree on that point – in the preface to his 1755 dictionary, Samuel Johnson wrote that "Words are seldom exactly synonymous" – but they do believe that for any pair of synonyms there are statements whose meanings (truth values) are not changed by using one of the words in place of the other. This weaker sense of synonymy – interchangeability in some contexts *salva veritate*, to use Leibniz's phrase – is a basic assumption underlying the art of lexicography.

So the anomaly is resolved, but it leaves "synonymy" in a dangerously ambiguous state. At least one philosopher (Goodman, 1949) has argued that the term should be dropped, that it is impossible to do better than to say that two words have a greater or lesser degree of "likeness of meaning". Psychologists concerned with this semantic relation have generally agreed; they have largely abandoned "synonymy" in favour of "similarity of meaning", "semantic distance", or more generally, "semantic similarity". This move not only avoids arguments about synonymy, but it introduces the idea of a continuous variable where before there seemed to be a presumption of dichotomy: that any two words either are synonyms or they are not synonyms. Armed with semantic similarity, it is not necessary to decide whether, say, *snake* and *serpent* are (exact) synonyms; it is enough to say that native speakers of English judge them to be highly similar in meaning. Indeed, native speakers are able to construct a series of pairs ordered for semantic similarity: *snake-serpent*, *snake-eel*, *snake-toad*, *snake-rat*, *snake-dog*, *snake-bush*, *snake-taxi*, *snake-autumn*, for example. Then, given such intermediate values of semantic similarity, it becomes possible to define variables like "typicality" in terms of the semantic similarity of a superordinate term to its various hyponyms (Rips, Shoben, & Smith, 1973). Typicality, of course, is but one of many uses that psychologists have found for semantic similarity.

Similarity has always played an important role in psychological theories, and semantic similarity has simply been adopted as a specific kind of similarity, often contrasted with formal or perceptual similarity. Over the years the collocation "semantic similarity" has crept into an increasing variety of psychological publications, sometimes used in the instructions given to subjects, sometimes referring to a variable assumed to influence subjects' responses, sometimes adopted as a criterion for scoring open-ended responses. A review of the extensive literature touching on semantic similarity would be an interesting contribution to the history of cognitive psychology, but it is beyond the scope of this paper. Here it is sufficient to make the point that most of these publications take it for granted that intelligent persons who know a language can reliably assess the semantic similarity of any two words they know how to use. This assumption has been supported by various studies. Henley (1969), for example, reported a reliability coefficient of 0.97 between the mean ratings of 435 pairs of animal names rated in two different sessions; Rubenstein and Goodenough (1965) reported equally high reliabilities for the average ratings of 65 pairs of nouns chosen to cover a wide range of similarities. But the presumptive validity of semantic similarity seems to be sufficient that theorists who use the concept seldom bother to cite precedents.

The result is that semantic similarity has become one of those ubiquitous and important variables, like familiarity or frequency of occurrence, that are often used to explain psychological phenomena, but are seldom seen as being in need of explanation. This degree of acceptance for semantic similarity is remarkable in view of the theoretical complexity of the judgement involved. Yet subjects accept instructions to judge similarity of meaning as if they understood immediately what is being requested, then make their judgements rapidly with no apparent difficulty.

The present paper takes semantic similarity as a dependent variable. The general idea is to consider semantic similarity as a function of the contexts in which words are used. The importance of context has long been recognised in discussions of synonymy, but context has received relatively little attention in discussions of semantic similarity.

A contextual approach to semantic similarity can be introduced as follows. The important concept in discussions of synonymy is the substitutability of one word for another without a change of truth values. Note that it is the invocation of truth values that gives synonymy its dichotomous character. When synonymy is generalised to semantic similarity, truth values are abandoned: in two-valued logics there are no intermediate truth values to go along with intermediate similarities. Substitutability need not be abandoned, however. Words with different but similar meanings can often be substituted for one another without changing a sentence into

## 4 MILLER AND CHARLES

something that no-one would be likely to use. The argument here will be that substitutability without loss of plausibility is an important factor underlying judgements of semantic similarity.

## CONTEXTUAL REPRESENTATIONS

What people know when they say that they know a word is not how to recite its dictionary definition – they know how to use it (when to produce it and how to understand it) in everyday discourse (Miller, 1986). Knowing how to use words is a basic component of knowing a language, and how that component is acquired is a central question for linguists and cognitive psychologists alike. The search for an answer can begin with the cogent assumption that people learn how to use words by observing how words are used. And because words are used together in phrases and sentences, this starting assumption directs attention immediately to the importance of context.

In discussions of vocabulary acquisition, "context" has a narrow and a broad sense. The narrow sense includes only linguistic context – the words uttered before and after the word in question. The broad sense also includes this linguistic context, but in addition it includes any non-linguistic information that a learner may have about the situation and the participants and the participants' communicative intentions. It is arguable that children do not acquire language when nothing more than the purely linguistic context is available to them. But, however that may be, it is certainly true that literate persons acquire many new words through reading, where nothing but a linguistic context is available. It is not unreasonable to assume, therefore, that people can learn how words are used on the basis of information provided by purely linguistic contexts.

If the narrow reading of "context" is assumed, then an important question is: What does a person learn from hearing or seeing a word used in a series of linguistic contexts? Presumably, an association is formed between an utterance (or inscription) and some representation of attributes common to all of those contexts. For the convenience of discussion, call this the word's contextual representation.

The *contextual representation* of a word is knowledge of how that word is used.

Note that a word's contextual representation differs from its phonetic, or morphophonological, representation. A possible metaphor would be that the phonetic representation of a word plays the same role as a name or address on a file; the file itself contains not only the word's contextual representation, but any other knowledge – orthographic, etymological, iconic, conceptual – that a person has learned about the word.

The choice of "contextual representation" may invite misunderstanding. "Representation in semantic memory" would be more conventional terminology. But "contextual representation" has two virtues: (1) it keeps the contextual origins of most lexical knowledge in plain view, and (2) it accommodates more than just semantic information about the word. A word's contextual representation is assumed to include syntactic, semantic, pragmatic, and stylistic conditions governing the use of that word.

So the question becomes: What is this contextual representation that results from hearing or seeing a word used in a variety of linguistic contexts? The simplest answer would follow a line suggested by Anderson and Bower (1973): analyse linguistic input into atomic propositions and store them directly in long-term memory. If all knowledge, including lexical knowledge, were stored in that form, a person's knowledge of word *A* would be given by all the contexts in which *A* had occurred. However, most theorists assume (as do the authors of this paper) that the cognitive representation of a word is some abstraction or generalisation derived from the contexts that have been encountered. That is to say, a word's contextual representation is not itself a linguistic context, but is an abstract cognitive structure that accumulates from encounters with the word in various (linguistic) contexts. The information that it contains characterises a class of contexts.

How the derivation is performed, how a contextual representation is derived from encounters with a series of specific contexts, is seldom specified. But some of the derivatives that seem to be required in order to account for language use have been described.

## Collocational Characterisation

Some part of knowing how to use a word is knowing the high-probability collocations into which it enters. From *absolute zero* to *zygomatic arch*, English is studded with strings of words that go together with far greater than chance probability. For example, *hermetically* only occurs in the context [ \_\_\_\_ sealed]; this context would, therefore, constitute part of the contextual representation stored with the word *hermetically*. Although the proper account of people's knowledge of collocations is still a matter for debate, such collocations are surely learned through association by contiguity after multiple encounters in different linguistic contexts.

Also to be included at this level of characterisation are such morphophonological rules as the one that turns the indefinite article *a* into *an* in front of words beginning with a vowel. Thus a rule something like

*a* → *an* / \_\_\_\_ # [vocalic]

## 6 MILLER AND CHARLES

would constitute part of the contextual representation stored with the indefinite article.

### Syntactic Characterisation

The most specific proposals have dealt with the syntactic component of a contextual representation. For example, most grammarians now assume that along with each word is stored a syntactic characterisation of the grammatical contexts in which that word occurs. (In recent years, grammarians have tended to increase the amount of syntactic information that they assume is stored with each word: Wasow, 1989.) For example, for purposes of syntax, *put* is assumed to be entered in the mental lexicon as a verb, and as requiring both a direct object and a prepositional phrase: [ \_\_\_\_ NP PP]. Thus, *He put the bread in the oven* is acceptable, whereas *He put the bread* or *He put in the oven* are not.

But knowing how to use *put* involves more than knowing the syntax of its contexts. Knowing how to use it involves additional information that makes it possible to block such further sentences as *He put the yell in the bank*, *He put Illinois in Chicago*, *He put the bread in the recognition*, etc. – sentences that are grammatical, but must be judged false, non-literal, or uninterpretable. Also associated with *put*, therefore, are semantic, pragmatic, stylistic, and general information to the effect that *put* is not used with those kinds (or pairs) of NPs and PPs. There is a strong presumption that such additional information is available from linguistic contexts in the fact that most words are learned from that source (Nagy & Herman, 1987; Sternberg & Powell, 1983).

### Semantic Characterisation

Lacking any theory of meaning comparable in detail or precision to syntactic theory, it is less clear how the semantic information in a word's contextual representation is to be characterised, but it is reasonably obvious that it is not represented in the paraphrase form made familiar by dictionary definitions. It is a common observation that people experience great difficulty in providing adequate definitions for words that they have no trouble using or understanding in everyday discourse. Putnam (1975) notes, for example, that a person may be able to use the word *gold* felicitously, yet be totally unable to provide a definition that would distinguish gold from other metals (that would make it possible to determine the truth or falsity of "This is gold" in every possible instance). Presumably, many encounters with a word in different contexts are sufficient for the acquisition (or recognition) of the lexical concept that the word is used to express, but this learning process involves a complex

interaction between what is already known or believed and any new information provided by contexts in which the word is used. The concepts that words express are not acquired in isolation, but are organised in semantic memory by lay theories about the domains in which they are used (Miller & Johnson-Laird, 1976; Murphy & Medin, 1985).

A rough sketch is possible, however. If we accept a simplified view of sentences as predicate–argument constructions, where verbs and adjectives serve as predicates and nouns or nounphrases serve as arguments, then sentences display to a learner the basic predicate–argument combinations used in the language. Context provides information about constraints on the arguments that a given predicate takes, and about constraints on the predicates that apply to a given argument. Such a system of constraints, when mapped onto a representation of the non-verbal world, provides a plausible account of an individual's semantic knowledge.

This approach to semantic and conceptual development has not been neglected. Keil (1979), for example, has reported an extensive investigation of vocabulary growth in terms of children's progressive mastery of such predicate–argument relations. It is necessary merely to add to Keil's account the claim that these predicate–argument relations are learned by observing them in the language of others. Berwick (1989) has described a computer program that learns word meanings from examples by taking advantage of various contextual constraints: the meaning of any word is represented by its position in a network that describes texts in which the word has occurred; a new word is learned by analogy – by matching its contextual network with the networks of familiar words to find the best agreement. In Berwick's view, therefore, judging the similarity of such semantic structures is the basic mechanism for learning new words from their contexts of use.

### Pragmatic and Stylistic Characterisations

Although reasonable proposals have been made concerning both the syntactic and the semantic information in a contextual representation, almost nothing is understood about the representation of pragmatic or stylistic information. Pragmatic constraints imposed by the linguistic context can be illustrated by antecedent/anaphor relations, or by question/answer relations. For example, the context *What colour is the \_\_\_\_ book?* self-destructs if a colour adjective is substituted into it. Stylistic constraints include the use of Latinate terms in formal or technical contexts and the use of Anglo-Saxon terms in informal or ribald contexts.

This catalogue of different kinds of information that can be acquired from a word's linguistic contexts and included in its contextual representation is intended to be suggestive, not definitive. The important point is that

## 8 MILLER AND CHARLES

this kind of information suffices to characterise abstractly a set of contexts in which the word is used. Someone who knows how to use a word knows whether or not any given context is a member of the set of contexts satisfying the word's contextual representation.

Sketchy as this account is, it is still possible to pose certain questions about contextual representations. A question of some importance concerns how the contextual representations for different words are related to one another. And the relation of special interest here, of course, is similarity: How well can the contextual representations of different words be discriminated from one another?

## A CONTEXTUAL HYPOTHESIS

An observation by the linguist Harris (1970, pp. 785–786) is relevant to the question of similarities between contextual representations:

If we consider words or morphemes *A* and *B* to be more different in meaning than *A* and *C*, then we will often find that the [contextual] distributions of *A* and *B* are more different than the [contextual] distributions of *A* and *C*. In other words, difference of meaning correlates with difference of [contextual] distribution.

If Harris is right – if this relation between similarity of meanings and similarity of contexts holds – it opens an interesting line for speculation. When people are given two words and asked to judge the similarity of their meanings, do they respond by judging the similarity of those words' contextual representations? That is to say, are two words judged to be similar in meaning to the extent that they are used in similar ways, i.e. are they used in similar (linguistic) contexts and so have similar contextual representations? These speculations can be summarised:

*Strong Contextual Hypothesis:* Two words are semantically similar to the extent that their contextual representations are similar.

This general line of thinking, emphasising the importance of linguistic context, has been explored both by linguists and psychologists. Joos (1950, p. 708), for example, claimed that "the linguist's meaning of a morpheme . . . is by definition the set of conditional probabilities of its occurrence in context with all other morphemes". Deese (1965), who defined the "associative meaning" of a word in terms of the distribution of responses it evokes on a word association test, phrased his version of the contextual hypothesis as follows: "the extent to which words share associative distributions is determined by the extent to which they share contexts in ordinary discourse" (p. 128). More recently, Cruse (1986) has written, "it

is assumed that the semantic properties of a lexical item are fully reflected in the appropriate aspects of the relations it contracts with actual and potential contexts" (p. 1), and "the meaning of a word is constituted by its contextual relations" (p. 16). Obviously, if meanings are contextual relations, the similarity of meanings and the similarity of contexts must covary.

Among psychologists, relations between semantic similarity and contexts of use have been discussed most extensively by advocates of associative theories learning. Most psychologists who have been interested in associative theories of learning have taken for granted a generalisation made explicit by Wicklund, Palermo, and Jenkins (1964, p. 418): "all associative habits tend to increase as a function of the frequency of occurrence of the contexts necessary for the formation of the associations". Because psychologists interested in learning theory have generally assumed that meanings are associative habits, the dependence of meanings on their contexts follows directly. Although this general approach led to several interesting observations about language, much of the work miscarried, apparently due to an over-emphasis on word associations as the fundamental empirical facts to be explained.

When the contextual hypothesis is formulated in this strong form, however, exceptions are quickly recognised. For example, someone who speaks both English and German would judge *department* and *Abteilung* to be similar in meaning, yet the linguistic contexts in which they occur are completely different. Even within one language there are exceptions: *department* and *departmental* are closely similar in meaning, yet they would almost never be used in the same linguistic contexts: nouns and adjectives are seldom interchangeable. A further qualification, therefore, is that the words whose similarity is being judged come from the same syntactic category. Morphological relations may play an important role in judgements of the semantic similarity of words from different syntactic categories, but little or nothing is known about it.

As it is possible to estimate the semantic similarity of words from different languages, or words from different syntactic categories – words that occur in very different contexts – the strong contextual hypothesis is apparently too strong. A weaker version might be stated:

*Weak Contextual Hypothesis:* The similarity of the contextual representations of two words contributes to the semantic similarity of those words.

In this version, other knowledge that a person has about the words can also contribute to judgements of their semantic similarity. The question that the weak contextual hypothesis leaves open, of course, concerns the relative importance of contextual representations in making such judgements. But

that is an empirical question that can be investigated by comparing measures of semantic similarity with measures of the similarity of contextual representations.

### MEASURES OF THE SIMILARITY OF SETS OF CONTEXTS

Semantic similarity is easily estimated by asking people to rate pairs of words with respect to their likeness of meaning. However, in order to test whether the semantic similarity of two words is correlated with the similarity of contexts in which those words occur, reliable estimates of contextual similarity are also necessary – and they have proven more difficult to obtain. Two different strategies have been tried, one based on co-occurrence and the other on substitutability, with rather different results.

Jenkins (1954), working within the associative tradition, first drew attention to the distinction between (1) syntagmatic word associations, which arise from the co-occurrence of words in discourse and are attributed to association by contiguity, and (2) paradigmatic word associations, which arise from the substitutability of words in discourse and are attributed to mediated association, i.e. to associations mediated by common contexts. Both kinds of associative learning were assumed to be necessary: association by contiguity in order to account for the learning of collocations; context-mediated association in order to account for the learning of syntactic categories (Ervin, 1961; 1963; Ervin-Tripp, 1970). It is noteworthy that the methods Jenkins (1954) proposed for determining whether a word association is primarily paradigmatic or syntagmatic correspond to the two methods that have been used to estimate contextual similarities.

Consider first a strategy based on *co-occurrence*: (1) list all the words that occur in a set of contexts of item A; (2) list all the words that occur in a comparable set of contexts of item B; then (3) calculate some normalised coefficient representing the proportion of words common to the two lists. The more likely it is that words co-occurring with A also co-occur with B, the more similar the two sets of contexts are judged to be.

A co-occurrence strategy was adopted by Rubenstein and Goodenough (1965) in a direct attack on the relation between semantic similarity and linguistic context. First, they obtained subjective ratings of semantic similarity for 65 pairs of English nouns. Then, as a measure of contextual similarity, they reasoned that similar contexts should contain the same words. So they obtained subject-generated sentences incorporating their target nouns and proceeded, for all 65 pairs, to count the number of other words that were common to these subject-generated contexts. Several measures of contextual similarity based on different amounts of context were tried: the general idea is to estimate the contextual “overlap” for

each noun pair by coefficients of correlation. The intersection of two sets of contexts  $C_1$  and  $C_2$  is given as a ratio of the number of words common to  $C_1$  and  $C_2$  divided by the larger of the absolute number of items in either  $C_1$  or  $C_2$ .

The results obtained were consistent with the contextual hypothesis, but Rubenstein and Goodenough (1965) were not satisfied with them. They concluded that:

it may be safely inferred that a pair of words is highly synonymous if their contexts show a relatively great amount of overlap. Inferences of degree of synonymy from lesser amounts of overlap . . . are uncertain since words of low to medium synonymy differ relatively little in overlap. (Rubenstein & Goodenough, 1965, p. 633).

In other words, this measure of contextual similarity, based on co-occurrence, confirmed the contextual hypothesis only for short distances in semantic space.

The second strategy is based on *substitutability*: (1) collect a set of sentences using item A; (2) collect a set of sentences using item B; then (3) delete A and B, shuffle the resulting contexts; and (4) challenge subjects to sort out which is which. The more contexts there are that will take either item, the more similar the two sets of contexts are judged to be.

This approach for estimating the similarity of sets of contexts has been called the method of sorting (Charles, 1988a; 1988b). A subject's task is to arrange sets of linguistic contexts for two (or more) words into groups of contexts all capable of accepting the same missing word; signal detection theory is then used to obtain an estimate of contextual discriminability. The method of sorting follows along the line of Leibniz's definition of synonymy in terms of interchangeability – if two words were perfect synonyms, it would be impossible to discriminate the contexts of one from the contexts of the other. The method of sorting departs from Leibniz's definition, however, by ignoring truth conditions: the only question asked of a subject is which word, when substituted into the context, would yield the most plausible sentence.

In studies using the method of sorting it was observed that this measure of contextual similarity, based on substitutability, varied directly with the semantic similarity of the target words. That is to say, the more similar in meaning the target words seemed to be, the more contexts there were that would take either term. For example, Charles (1988a) found that sets of sentential contexts were more similar for the semantically related terms *hotel-building* than for the controls *car-ball*.

In order to test whether or not this relation was accidental or systematic, it was decided to repeat Rubenstein and Goodenough's (1965) experiment,

## 12 MILLER AND CHARLES

but this time using a measure of the similarity of sets of contexts based on substitutability rather than co-occurrence. Of particular interest was the relation of semantic similarity to contextual similarity at low levels of contextual similarity (high levels of contextual discriminability), because that was where estimates based on co-occurrence had been least satisfactory.

### EXPERIMENT 1: RATINGS FOR SEMANTIC SIMILARITY

Experiment 1 was designed to determine whether a new group of subjects would agree with the subjects used by Rubenstein and Goodenough (1965) in judging semantic similarity.

#### Subjects

A total of 38 students at the State University of New York at Oswego participated as subjects. All of them were native speakers of English.

#### Materials

A subset of 30 noun pairs from the original list of 65 studied by Rubenstein and Goodenough (1965) was used; 10 were selected from the high level (between 3 and 4 on a scale from 0 to 4), 10 from the intermediate level (between 1 and 3), and 10 from the low level (0 to 1) of semantic similarity. The 30 test pairs were then printed on two separate sheets of paper.

#### Instructions and Procedure

Each subject was tested individually. The subjects were told to judge similarity of meaning. As examples of degrees of synonymy, the pairs *gem-jewel*, *bird-cock*, and *autograph-shore*, which were rated high, intermediate, and low respectively by Rubenstein and Goodenough (1965), were shown to each subject. A subject was then presented with the two sheets of paper on which 30 noun pairs appeared and was instructed to examine each pair closely and then to rate it on a 5-point scale from 0 to 4, where 0 represents no similarity of meaning and 4 perfect synonymy. The ordering of the pairs was randomly determined for each subject. The subjects were free to rate and re-rate the pairs for as long as they chose.

#### Results and Discussion

The mean ratings for the 30 pairs are reported in Table 1 along with those obtained by Rubenstein and Goodenough (1965). The two sets of ratings are in good correspondence: the Pearson product-moment correlation coef-

TABLE 1  
Ratings of Semantic Similarity for 30 Pairs of Nouns

Noun Pair	Oswego	R & G
<i>car-automobile</i>	3.92	3.92
<i>gem-jewel</i>	3.84	3.94
<i>journey-voyage</i>	3.84	3.58
<i>boy-lad</i>	3.76	3.82
<i>coast-shore</i>	3.70	3.60
<i>asylum-madhouse</i>	3.61	3.04
<i>magician-wizard</i>	3.50	3.21
<i>midday-noon</i>	3.42	3.94
<i>furnace-stove</i>	3.11	3.11
<i>food-fruit</i>	3.08	2.69
<i>bird-cock</i>	3.05	2.63
<i>bird-crane</i>	2.97	2.63
<i>tool-implement</i>	2.95	3.66
<i>brother-monk</i>	2.82	2.74
<i>lad-brother</i>	1.66	2.41
<i>crane-implement</i>	1.68	2.37
<i>journey-car</i>	1.16	1.55
<i>monk-oracle</i>	1.10	0.91
<i>cemetery-woodland</i>	0.95	1.18
<i>food-rooster</i>	0.89	1.09
<i>coast-hill</i>	0.87	1.26
<i>forest-graveyard</i>	0.84	1.00
<i>shore-woodland</i>	0.63	0.90
<i>monk-slave</i>	0.55	0.57
<i>coast-forest</i>	0.42	0.85
<i>lad-wizard</i>	0.42	0.99
<i>chord-smile</i>	0.13	0.02
<i>glass-magician</i>	0.11	0.44
<i>rooster-voyage</i>	0.08	0.04
<i>noon-string</i>	0.08	0.04

Note: Mean ratings on a scale from 0 to 5 by 38 subjects in Experiment 1 (Oswego) compared with mean ratings reported by Rubenstein & Goodenough (R & G).

ficient is 0.97, significant at the 0.01 level. People are not only able to agree reasonably well about the semantic distances between concepts, but their average estimates remain remarkably stable over more than 25 years.

### EXPERIMENT 2: SORTING SENTENTIAL CONTEXTS

The method of sorting contexts was used to estimate contextual similarity for 6 of the 30 pairs of nouns that were rated for synonymy in Experiment 1. Two pairs represented each level of synonymy: the high-level pairs were

## 14 MILLER AND CHARLES

*car-automobile* and *coast-shore*; the intermediate-level pairs were *fruit-food* and *car-journey*; the low-level pairs were *monk-slave* and *noon-string*.

The pair *fruit-food* was included despite being rated higher in Experiment 1 than in the earlier study. The decision to include it was influenced by the fact that *fruit* is a hyponym of *food*. Charles (1988a) found that, in the case of *hotel-building*, hyponymy led to poor contextual discriminability and showed a strong response bias favouring the superordinate term, *building*. *Fruit-food* was included in order to see whether that earlier result would replicate.

**Subjects**

A group of 24 undergraduate students at the State University of New York at Oswego, not including anyone who participated in Experiment 1, were tested. All of them were native speakers of English.

**Materials**

Contexts for the six pairs of nouns were collected by a computer search of the Brown Corpus (Francis and Kučera, 1982). That is to say, the sentential contexts used in this experiment were derived from sentences composed by authors whose communicative intentions were unrelated to the present experiment, and so represent how these nouns are used unconsciously in writing. Most of the words occur in 25 (or more) sentences in the Brown Corpus, but only 22 are available for *monk-slave* and *noon-string*. The Brown Corpus includes enough sentences using *car* that two different sets of *car*-contexts were available, one set for the *car-automobile* comparison and another for *car-journey*. The sentences were converted to contexts by replacing the target words, both singular and plural, by a dash.

The decision to study only 6, rather than the entire set of 30 noun pairs, was influenced both by time considerations and by the availability of suitable sentences in the Brown Corpus. The following contexts illustrate the kinds of stimuli that subjects were asked to sort. For the pair *car-automobile*, respectively:

I found a parking place half a block away, sat in the \_\_\_\_\_ and waited.  
The bombs are as harmless as an \_\_\_\_\_ in a garage.

Note that *an* in the second context could not be followed by *car*. For the pair *car-journey*, respectively:

He replaced the flashlight where it had been stowed, got into his own \_\_\_\_\_ and backed it out of the garage.

They began the \_\_\_\_\_ through their own mine fields.

For the pair *monk-slave*, respectively:

A second tale shows still more clearly the kind of powers a truly spiritual \_\_\_\_\_ could possess.

Besides he owns 300 \_\_\_\_\_.

And for the pair *noon-string*, respectively:

The dinner hour was 12.00 \_\_\_\_\_.

Nerves tight as a bow \_\_\_\_\_, he paused to gather his wits.

**Method**

A within-subjects design was used. Each subject was tested individually and sorted contexts for one pair of nouns at a time, moving on to the next pair after completing each sort. The subjects were told in advance what the target words were: the labels under which the contexts were to be sorted were taped to the desk at which the subjects worked. The subjects were asked to sort the contexts into two groups, and to include in each group contexts that would accept the same word. The order of the sortings was counterbalanced by synonymy.

**Results**

The six pairs of nouns are listed in decreasing order of semantic similarity (as determined in Experiment 1) in Table 2, which summarises the sorting data. For example, *car* contexts were sorted 600 times (25 contexts × 24 subjects); 524 times they were assigned to the label "car" and 76 times they were assigned to the label "automobile". The rest of the table is read in the same manner.

According to Table 2, contexts for both pairs of semantically similar nouns, *car-automobile* and *coast-shore*, were the most frequently confused for one another. The main difference between the results for these two pairs was in the response bias. *Car* was strongly favoured over *automobile*, whereas contexts were assigned to *shore* only slightly more frequently than to *coast*.

The results for *food-fruit* show no significant response bias in favour of the generic term over its hyponym. Apparently, the bias in favour of *building* over *hotel* observed in a previous study is peculiar to that pair of words and is not a general characteristic of pairs related to hyponymy.

The contexts for *food-fruit* were more frequently confused than were those of *car-journey*, which, in turn, were more frequently confused than were the *noon-string* contexts. Thus, the relation between substitutability

TABLE 2  
Sorting Data for Contexts Drawn from the Brown  
Corpus

Stimulus	Response		Sum
	Car	Automobile	
Car	524	76	600
	258	342	600
	782	418	1200
Coast	440	160	600
	150	450	600
	590	610	1200
Food	515	85	600
	89	511	600
	604	596	1200
Journey	576	24	600
	14	586	600
	590	610	1200
Monk	469	59	528
	44	484	528
	513	543	1056
Noon	522	6	528
	2	526	528
	524	532	1056

Note: Cell entries show the number of times the contexts for a given word were sorted correctly and incorrectly.

and similarity of meaning is straightforward, with one exception, *monk-slave*, which, though judged rather different in meaning on the rating scale, were substitutable for one another relatively frequently.

#### Signal Detection Theory

The data in Table 2 are conveniently summarised by the use of signal detection theory (Egan, 1975). Consider the pair *car-automobile*, for example. If *car* is taken as the signal to be detected, then the probability of

TABLE 3  
Discriminability of Contexts Drawn from the Brown Corpus

Stimuli	<i>d'</i>	Similarity Rating
<i>Car-automobile</i>	1.22	3.92
<i>Coast-shore</i>	1.30	3.70
<i>Food-fruit</i>	2.11	3.08
<i>Journey-car</i>	3.74	1.16
<i>Monk-slave</i>	2.60	0.55
<i>Noon-string</i>	4.95	0.08

Note: The estimate of contextual discriminability, *d'*, is shown with the estimated semantic similarity for all six pairs of nouns.

a hit,  $P(\text{"car"}|\text{car})$  – the probability of sorting a context under the label “car” when it was a *car* context – can be estimated as  $524/600 = 0.873$ ; the probability of a false alarm,  $P(\text{"car"}|\text{automobile})$ , can be estimated as  $258/600 = 0.430$ . These probabilities correspond to a *d'* of 1.22. [Alternatively, if *automobile* is taken as the signal, then the probability of a hit,  $P(\text{"automobile"}|\text{automobile})$ , is  $342/600 = 0.570$ ; the probability of a false alarm,  $P(\text{"automobile"}|\text{car})$ , is  $76/600 = 0.127$ ; and these probabilities also yield a *d'* of 1.22.] Similar calculations of *d'* for all six pairs of nouns are shown in Table 3.

#### Discussion

Signal detection analysis confirms the impression based on inspection of the data in Table 2: an inverse linear relation between semantic and contextual similarity is observed for five of the six pairs, extending over the entire range that was examined. It appears, therefore, that a measure of contextual similarity based on substitutability gives better predictions of semantic similarity than did Rubenstein and Goodenough's (1965) measure of contextual similarity based on co-occurrence.

An exception to the general trend is *monk-slave*, whose semantic similarity was rated lower, relative to the other five pairs, than the estimate of contextual similarity would predict. This deviation is shown graphically in Fig. 1, where the mean value of semantic similarity is plotted as a function of *d'*, the measure of contextual discriminability; *monk-slave* is the filled square that falls below the line fitting the other five points.

It is unlikely that this pair deviated because of faulty ratings; the Oswego subjects agreed with Rubenstein and Goodenough's (1965) subjects in assigning *monk-slave* a low rating for semantic similarity. If this exception

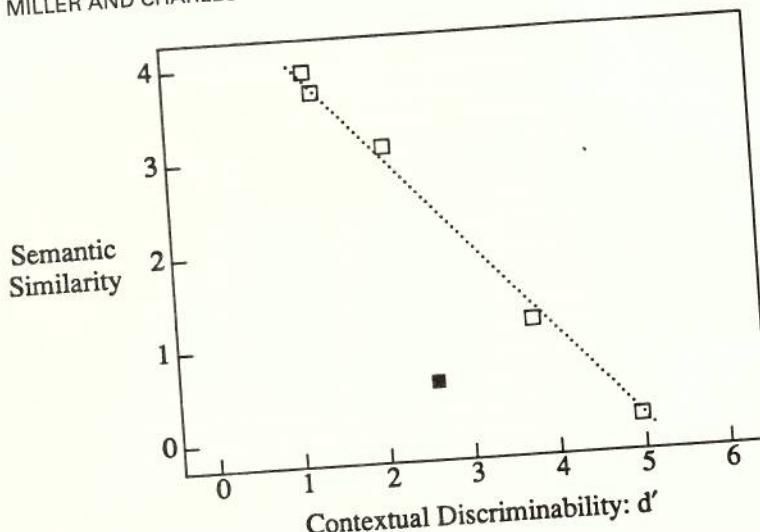


FIG. 1 Semantic similarity plotted as a function of contextual similarity (based on contexts drawn from the Brown Corpus).

is due to a chance deviation from expectation, therefore, it is probably attributable to the sample of sentences provided by the Brown Corpus. In view of this possibility, a new sample of sentences was obtained, this time by the technique that Rubenstein and Goodenough used: the subjects were instructed to write sentences that included the target words.

### EXPERIMENT 3: SORTING SUBJECT-GENERATED SENTENTIAL CONTEXTS

Experiment 3 replicated Experiment 2, using the same six pairs of nouns, but different sets of sentential contexts. The replication was intended to test the reliability of the results obtained in Experiment 2; new sentences were chosen to explore possible differences between sentences that occur in published texts and sentences that are generated in isolation simply for the purposes of an experiment.

#### Subjects

Another group of 24 subjects performed the sorting task. All were native speakers of English, all were undergraduates at the State University of New York at Oswego, and none had participated previously in Experiments 1 or 2.

#### Materials

Two groups of 12 subjects each were enlisted to generate a corpus of sentences containing the 12 words to be compared. One group received a list containing six nouns, one from each of the six pairs; the other group received a list of the other six nouns. They were asked to write two sentences that included each word appearing on their list. The sentences were to be between 10 and 15 words long. No restriction was placed on the time allowed to complete the task. Then 25 of these subject-generated sentences were selected at random to provide contextual stimuli for Experiment 3.

The following contexts illustrate the kinds of stimuli that were produced in this manner. For the pair *car* and *automobile*, respectively:

I drove my \_\_\_\_ back up to school last weekend.

Henry Ford is the inventor of the \_\_\_\_.

For the pair *car-journey*, respectively:

The married couple just bought their first \_\_\_\_.

I would love to take a long \_\_\_\_ across the United States.

For the pair *monk-slave*, respectively:

A \_\_\_\_ dedicates his life to God.

It makes me mad to think about all the \_\_\_\_ there were in the south.

#### Method

The method of sorting was again used to estimate contextual similarity in a manner consistent with Experiment 1.

#### Results

The data for subject-generated contexts is summarised in Table 4, which parallels Table 2. Table 4 is remarkably similar to Table 2. The relation of semantic similarity to contextual discriminability is the same in both sets of data: *car-automobile* and *shore-coast* are the most similar semantically and their contexts are still the hardest to discriminate; *noon-string* are semantically unrelated and their contexts are again the easiest to discriminate. The exception is *monk-slave*, which was also an exception in Experiment 2. Even the strong response bias in favour of the term *car* in the pair *car-automobile* is the same.

TABLE 4  
Sorting Data for Contexts Based on Subject-generated Contexts

Stimulus	Response		Sum
	Car	Automobile	
Car	511	89	600
	324	276	600
	835	365	1200
Automobile	Coast	Shore	
	398	202	600
	127	473	1200
Sum	525	675	
	Food	Fruit	
	545	55	600
Food	40	560	600
	585	615	1200
	Journey	Car	
Journey	596	4	600
	2	598	
	598	602	1200
Car	Monk	Slave	
	575	25	600
	10	590	
Slave	585	615	1200
	Noon	String	
	599	1	600
Noon	0	600	600
	599	601	1200
	Sum		

Note: Cell entries show the number of times the contexts for a given word were sorted correctly and incorrectly.

### Discussion

The results for the two corpora are similar, but not identical. The data in Table 4 were analysed by signal detection theory and the resulting  $d'$  values are shown in Table 5 along with the ratings for semantic similarity. Table 5, for subject-generated contexts, is comparable to Table 3 for sentential contexts taken from the Brown Corpus.

TABLE 5  
Contextual Discriminability for Subject-generated Contexts

Stimuli	$d'$	Similarity Rating
Car-automobile	1.14	3.92
Coast-shore	1.22	3.70
Food-fruit	2.82	3.08
Journey-car	5.19	1.16
Monk-slave	3.85	0.55
Noon-string	> 5.98	0.08

Note: The estimate of contextual discriminability,  $d'$ , is shown with the estimated semantic similarity for all six pairs of nouns.

That is a hint in these data that subject-generated contexts reflect more directly the underlying structure of semantic memory for these words. In particular, the range of  $d'$  is greater for data obtained with the subject-generated corpus than for data obtained with sentences taken from the Brown Corpus, and the slope of the function relating semantic similarity to contextual discriminability is correspondingly less steep. This difference is apparent when the functions in Figs 1 and 2 are compared.

It is also apparent from Fig. 2 that the semantic similarity of *monk* and *slave* is again less than expected on the basis of contextual similarity. The same pair (shown by the filled square) is deviant for both sets of contexts; it is highly unlikely that some unforeseen factor led to the independent selection of two sets of contexts adventitiously less discriminable (more similar) than average. The results for *monk-slave* have some special significance.

In discussing the contextual hypothesis, it was noted that words to be compared should be chosen from the same language and from the same syntactic category. Those considerations led to a weakened version of the contextual hypothesis. Perhaps it is also important to choose words from the same semantic field. The most obvious thing distinguishing *monk-slave* from the other five pairs is that monks and slaves are people. Nouns denoting animate things in general, and people in particular, have special privileges of occurrence. Presumably, person terms are treated differently because language is spoken by people, to people, and (often) about people. Thus, for example, the vast majority of verbs will accept person nouns or nounphrases as their grammatical subject, so there are likely to be relatively many contextual opportunities for the plausible substitution of semantically dissimilar person terms. Thus, in such contexts as "Should

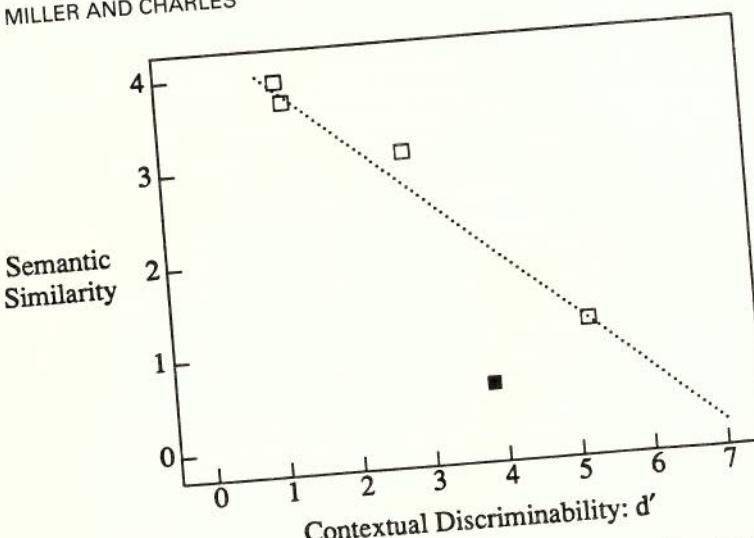


FIG. 2 Semantic similarity plotted as a function of contextual similarity (based on contexts generated by subjects).

any \_\_\_\_ change his mind and request to leave earlier, . . ." (from the Brown Corpus) or "\_\_\_\_ live a life of poverty" (from the subject-generated corpus), either *monk* or *slave* (or any one of many other person terms) is perfectly plausible. Only a few such contexts are needed in order to reduce  $d'$  significantly.

The possibility should be considered, therefore, that the relation between semantic similarity and contextual similarity is different in different semantic fields. That is to say, if *monk-slave* were compared with other pairs of person terms, a regular function might well emerge, but with parameters different from those obtained for comparisons of terms denoting inanimate things. This possibility is incompatible with the strong contextual hypothesis. In terms of the weak contextual hypothesis, however, it amounts to a claim that the similarity of contextual representations is less important (is assigned less weight) for pairs of person terms than for pairs of inanimate terms. Which suggests a speculation: If these comparisons were carried out extensively, the existence of different weights might provide a criterion for defining membership in a semantic field.

### GENERAL DISCUSSION

The weak contextual hypothesis states that contextual similarity contributes to semantic similarity. The present results are consistent with that hypothesis. Thus, the present results, using an estimate of contextual

similarity based on substitutability, provide much stronger support for a contextual hypothesis than do the data of Rubenstein and Goodenough (1965), whose estimates of contextual similarity were based on co-occurrences.

Three aspects of these findings merit discussion: first, the differences between the two measures of contextual similarity; secondly, the implications of the contextual hypothesis; finally, a word must be said about adjectives, which pose a critical test for contextual hypotheses.

### Measures of Contextual Similarity

In adopting a measure of contextual similarity based on co-occurrences, Rubenstein and Goodenough (1965) were simply doing what everyone thinks of first. Contexts are collections of words; to determine how similar two contexts are, count the number of words that appear in both collections.

Psychologists are probably most familiar with this strategy in Deese's (1965) use of an "intersection coefficient" to estimate the overlap between two distributions of word associations. He found that intersection coefficients for word associations to nouns are relatively small, although the trend supported his expectation that shared associates signal similar meanings. One major finding of Deese's (1964) studies, however, was a robust intersection coefficient for antonymous adjectives: *weak*, for example, is the most common response to the stimulus *strong*, and *strong* is the most common response to *weak*. This reciprocity inflates the intersection coefficient and gives a clear signal that two adjectives are related by antonymy. Because this success with a co-occurrence measure is unusual, it will be discussed further below.

An advantage of measures based on co-occurrence is that they are easily calculated with the help of modern computers. For example, Lesk (1986) used such a measure in a computer program intended to disambiguate words in written texts. He counted the number of words shared between (1) a sentence that contained an ambiguous word and (2) alternative definitions of that word in a machine-related dictionary. The sense whose definition had the greatest overlap with the text sentence (the most words in common) was chosen as the appropriate interpretation of the word in that particular context. Lesk found that the result was often correct, but not reliable enough for practical applications.

The problem with co-occurrence measures is not merely that they dismember the contexts they are supposed to represent. A more serious problem is that they do not approach these tasks the way people do — whatever a word's contextual representation may be, it is certainly not a collection of other words. If the argument advanced here is correct,

people's knowledge of how to use a word is organised to enable them to recognise rapidly the contexts it goes into. Consequently, measures of contextual similarity based on substitutability come closer to the desired goal. But the disadvantage of measures based on substitutability is that there is no quick and easy computer algorithm for calculating them.

### Implications of the Contextual Hypothesis

If these results are accepted as support for a contextual hypothesis – as supporting the notion that judgements of the semantic similarity of two words are strongly influenced by the similarity of those words' contextual representations – then they are also consistent with the notion that knowing a word is knowing how to use it, i.e. that encounters with a word in various contexts lead eventually to the construction of a contextual representation that characterises the contexts in which the word is used.

The notion of contextual representations for words is unfortunately vague, but the general idea is to consolidate various kinds of information about a word's contexts into a single representation that characterises those contexts. This proposal is here shown to be consistent with people's ability to judge semantic similarity; it is not inconsistent with what is known of the "top-down" comprehension process in reading or listening to discourse, or with what little is known about the processes involved in learning new words from context. Moreover, the fact that the relevant information in a word's contextual representation is information needed to characterise the contexts in which the word occurs, provides a theorist with useful constraints on the way this information should be characterised. But the important point to remember is that contextual representations are highly complex knowledge structures, and that matching them against a particular input structure is more like reasoning by analogy than like an associative reflex.

Finally, nothing that has been said about contextual representations should be taken to imply that this is the only kind of information that is associated with words. For example, the fact that *duck* and *bird* are judged less similar than *robin* and *bird* (Rips et al., 1973) would lead to a prediction that the contexts of *duck* and *bird* are less similar than the contexts of *robin* and *bird*, but it says nothing about the possibility that, quite independently, imagery is stored with *duck* and *robin*, and taxonomic knowledge is stored with *duck*, *robin*, and *bird*. It seems reasonable to suppose that the referents of a word have their own perceptual and conceptual representations that are related in ways having little or nothing to do with the relations discussed here between contextual representations.

### A Critical Case: Antonymous Adjectives

The case for the weak contextual hypothesis and the reality of contextual representations has been argued so far with only a bare mention of the most likely source of counter-examples, namely, antonymous adjectives. The problem is easily stated. Antonyms have contrasting meanings – not just zero semantic similarity, but negative similarity, if that is possible. Yet they seem to be freely substitutable for one another: If the referent of a head noun has a particular attribute, the noun could take either polar value of that attribute. For example, if it has size, it is either large or small; if it has weight, it is either light or heavy; and so on. Here, apparently, is a combination that the contextual hypothesis denies: high substitutability associated with low semantic similarity.

The substitutability of antonyms, however, is not what it seems. In principle, antonyms should be substitutable for one another; it is easy to think of sentences in which either antonym is as plausible as the other. But what possibly could happen is not what usually does happen. The fact is that the sentences containing adjectives that people actually utter or write do not accept antonyms interchangeably; one antonym is almost always more plausible than the other. Charles and Miller (1989) used sentences from the Brown Corpus (Francis & Kučera, 1982) to obtain sentential contexts for the semantically related adjectives *strong*, *weak*, *powerful*, and *faint*, and also for *public*, *private*, *open*, and *secret*. The discriminability of these contexts was then estimated by the method of sorting. In both cases, the contexts for synonymous pairs (*strong-powerful*, *weak-faint*, *public-open*, *private-secret*) were more similar (less discriminable) than were the contexts for the corresponding antonymous pairs (*strong-weak*, *powerful-faint*, *public-private*, *open-secret*). Surprisingly, then, antonyms are not an exception to a contextual hypothesis. (The unexpectedness of this result led to the more systematic test of the contextual hypothesis that is reported here.)

This leads us to a second question: If antonyms are not substitutable for one another in everyday texts and discourse, how do they come to be so strongly associated? Deese (1964) demonstrated systematically what other workers had observed incidentally with word associations, namely, that the strongest associate to an adjective is its antonym (if it has one). If this strong association does not arise as a consequence of the substitutability of antonyms, then it must result from their co-occurrence. Charles and Miller (1989) explored this possibility. They report results for *big*, *little*, *large*, and *small* based on the 54 717 sentences in the Brown Corpus (Francis & Kučera, 1982) showing that both *big-little* and *large-small* co-occur in sentences far more frequently than would be predicted by chance, whereas

*big-small* and *large-little* co-occur in sentences only at the chance level. Justeson and Katz (1989) extended this analysis to all of the high-frequency adjectives and their antonyms in the Brown Corpus and found that high co-occurrence rates do characterise all antonymous adjective pairs. Contrary to what many psycholinguists would expect, the strong associations between antonymous pairs cannot be attributed to their interchangeability in contexts, but could easily be acquired through association by contiguity.

So the situation is this: Antonyms tend to occur together in the same sentences, so if a measure of contextual similarity based on co-occurrence is used, a true counter-example to the contextual hypothesis is obtained: high contextual similarity coupled with low semantic similarity. This situation had been noted before, of course, but was dismissed with a rationalisation that antonyms may not be as different in meaning as they seem. The alternative conclusion, more consistent with the present account, is that co-occurrence does not provide a satisfactory basis for estimates of contextual similarity.

### CONCLUSION

Several linguists and psychologists have claimed to see a close connection between the meanings of words and the contexts in which they are used. A specific version of that general claim has been elaborated here in terms of similarity of meanings and similarity of contexts.

It has been found that when the similarity of sets of contexts for two words is estimated by a technique based on substitutability, the results predict mean ratings of the semantic similarity of those words. This relation is interpreted to mean that as people learn a word they acquire a mental representation of the contexts in which the word occurs, a representation that includes all of the syntactic, semantic, pragmatic, and stylistic information required to use the word appropriately. This contextual representation is called on when subjects are asked (1) to determine whether substituting a particular word into a particular context would yield a plausible sentence, or (2) to give subjective estimates of that word's semantic similarity to other words.

Many different psychological theories and experiments have invoked the concept of semantic similarity. If this variable depends on contextual representations as proposed here, some of those theories and experiments may benefit from reinterpretation.

Manuscript received September 1989  
Revised manuscript received July 1990

### REFERENCES

- Anderson, J. R. & Bower, G. H. (1973). *Human associative memory*. Washington, D.C.: V. H. Winston.
- Berwick, R. C. (1989). Learning word meanings from examples. In D. Waltz (Ed.), *Semantic structures: Advances in natural language processing*, pp. 89-124. Hillsdale, N.J.: Lawrence Erlbaum Associates Inc.
- Charles, W. G. (1988a). *Categorization of contextual information*. Ph.D. dissertation. Department of Psychology, Princeton University.
- Charles, W. G. (1988b). The categorization of sentential contexts. *Journal of Psycholinguistic Research*, 17, 403-411.
- Charles, W. G. & Miller, G. A. (1989). Contexts of antonymous adjectives. *Applied Psycholinguistics*, 10, 357-375.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
- Deese, J. (1964). The associative structure of some common English adjectives. *Journal of Verbal Learning and Verbal Behavior*, 3, 347-357.
- Deese, J. (1965). *The structure of associations in language and thought*. Baltimore, Md.: Johns Hopkins Press.
- Egan, J. P. (1975). *Signal detection theory and ROC analysis*. London: Academic Press.
- Ervin, S. M. (1961). Changes with age in the verbal determinants of word-association. *American Journal of Psychology*, 74, 361-372.
- Ervin, S. M. (1963). Correlates of associative frequency. *Journal of Verbal Learning and Verbal Behavior*, 1, 422-431.
- Ervin-Tripp, S. M. (1970). Substitution, context, and association. In L. Postman & G. Keppel (Eds), *Norms of word association*, pp. 383-467. London: Academic Press.
- Francis, W. N. & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.
- Goodman, N. (1949). On likeness of meaning. *Analysis*, 10, 1-7.
- Harris, Z. S. (1970). *Papers in structural and transformational linguistics*. Dordrecht: Reidel.
- Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior*, 8, 176-184.
- Jenkins, J. J. (1954). Transitional organization: Association techniques. In C. E. Osgood & T. A. Sebeok (Eds), *Psycholinguistics: A survey of theory and research problems*. *Journal of Abnormal and Social Psychology*, 52, 112-118 (suppl.).
- Joos, M. (1950). Description of language design. *Journal of the Acoustical Society of America*, 22, 701-708.
- Justeson, J. & Katz, S. (1989). *Co-occurrences of antonymous adjectives and their contexts*. Yorktown Heights, N.Y.: IBM Research Report.
- Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, Mass.: Harvard University Press.
- Lesk, M. E. (1986). Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice-cream cone. In *Proceedings of the 1986 ACM SIGDOC Conference*, Toronto, Canada.
- Lyons, J. (1977). *Semantics*, Vols 1 and 2. Cambridge: Cambridge University Press.
- Miller, G. A. (1986). Dictionaries in the mind. *Language and Cognitive Processes*, 1, 171-185.
- Miller, G. A. & Johnson-Laird, P. N. (1976). *Language and perception*. Cambridge, Mass.: Harvard University Press.
- Murphy, G. L. & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.
- Nagy, W. E. & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24, 237-270.

- Putnam, H. (1975). The meaning of "meaning". In K. Gunderson (Ed.), *Language, mind, and knowledge*, pp. 131-193. Minneapolis: University of Minnesota Press.
- Quine, W. V. O. (1953). Two dogmas of empiricism. Reprinted in Quine, W. V. O., *From a logical point of view*, pp. 20-46. Cambridge, Mass.: Harvard University Press.
- Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.
- Rubenstein, H. & Goodenough, J. B. (1965). Contextual correlates of synonymy. *Computational Linguistics*, 8, 627-633.
- Sternberg, R. & Powell, J. (1983). Comprehending verbal comprehension. *American Psychologist*, 38, 878-893.
- Wasow, T. (1989). Grammatical theory. In M. I. Posner (Ed.), *Foundations of cognitive science*, pp. 161-205. Cambridge, Mass.: MIT Press.
- Wicklund, D. A., Palermo, D. S., & Jenkins, J. J. (1964). The effects of associative strength and response hierarchy on paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 3, 413-420.