Applied Psycholinguistics **21** (2000), 505–524
Printed in the United States of America

Contextual correlates of meaning

WALTER G. CHARLES
Delaware State University

ADDRESS FOR CORRESPONDENCE
Walter G. Charles, Delaware State University, Psychology Department, 1200 North Dupont Highway, Dover, DE 19901-2277. Email: wcharles@dsc.edu

ABSTRACT
The relation between similarity and dissimilarity of meaning and similarity of context was analyzed for synonymous nouns. New semantic similarity and dissimilarity rating tests with an empirically determined series of linguistic anchors and conventional, arbitrarily anchored semantic similarity ratings were compared. Contextual similarity was elicited by a sorting test based on substitution and yielding d-primes. The study found reliable correlations between the d-primes and the different ratings for semantic similarity and dissimilarity of the synonymous nouns across a wide continuum of meaning. The data strongly supported a contextual hypothesis of meaning. The data endorsed the claim that people abstract a contextual representation from experiencing the multiple natural linguistic contexts of a word. Semantic similarity and dissimilarity rating formats with an empirically chosen series of linguistic anchors and a sorting test of contextual similarity yielded stronger support for a contextual hypothesis than did alternative methods of eliciting lexical and contextual similarity.

Synonymy is an important concept in linguistics and philosophy, but one which is not easy to define. According to a statement usually attributed to Gottfried Leibniz, two words are synonymous if they are interchangeable in statements without a change in the truth value of the statements where the substitution occurs. This definition of synonymy in terms of truth value emphasizes its dichotomous rather than its continuous nature: words are either synonymous or they are not. However, philosophical semanticists (Quine, 1953), lexicographers (Johnson, 1755, in Miller & Charles, 1991), and linguists (Cruse, 1986), while recognizing synonymy as a basic semantic relation, have rejected a truth value notion of synonymy. They suspect that, for all synonyms, there are statements whose meanings (or truth values) are changed by the substitution of one word for another. A weaker definition of synonymy attributed to Leibniz makes it relative to context. That is, synonyms are words that are interchangeable in some contexts. This weaker definition of synonymy is a basic premise in the scientific study of meaning.

Psychologists have made considerable use of the concept of similarity. They also have abandoned the concept of "synonymy" in favor of a weaker form of the concept: namely, "similarity of meaning," "semantic distance," or more generally "semantic similarity and dissimilarity." Similarity and dissimilarity play a critical role in categorization (Rosch, 1973; Smith, Shobin, & Rips, 1974), object recognition (Tversky, 1977), word acquisition (Carey, 1985),

© 2000 Cambridge University Press 0142-7164/00 $9.50

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

memory organization (Tulving & Thompson, 1975), problem solving and transfer (Glick & Holyoak, 1983), and the understanding of connotative meaning (Osgood, Suci, & Tannenbaum, 1957). Few areas of scientific psychology are immune from the acclaimed role that similarity plays in theory construction. A cogent history of the concept of similarity in psychology is found in Gentner and Rattermann (1991).

Semantic similarity and dissimilarity constitute a special type of similarity. Semantic similarity and dissimilarity deal with the literal meanings of words and the semantic domains in which lexical concepts are organized. Literal similarity and dissimilarity are often contrasted with iconic, analogical, and connotative similarity. The concepts of semantic similarity and dissimilarity are so familiar that native speakers accept instructions to make such judgments without needing further clarification, and they execute these judgments reliably. Indeed, judgments of semantic similarity and dissimilarity are often closely associated (Henley, 1969).

Semantic similarity of nouns and adjectives and, to a lesser extent, verbs and adverbs has been the focus of most psychological studies. Studies have shown that noun meanings primarily take advantage of hyponymic relations between concepts. Rosch (1973) argued that everyday semantic categories of noun concepts are cognitively organized in decreasing levels of generality, from superordinate to basic to subordinate meanings. Basic concepts are psychologically salient and maximize cognitive economy in representing category information. Prototypicality, family resemblance, and similarity and dissimilarity are other semantic relations that are critical to the internal structure of such categories.

In contrast, Deese (1964) and others proposed that bipolarity and semantic similarity are critical to the cognitive structure of adjectives. According to Charles and Miller (1989) and Gross, Fisher, and Miller (1989), associative bipolarity is critical to the representation of adjectival antonyms (*strong–weak*), whereas semantic opposition (*strong–faint*) and semantic similarity (*strong–powerful*) of adjectives are conceptual dimensions. However, Murphy and Andrew (1993) argued that conceptual features are critical to the representations of adjectival antonymy, semantic opposition, and semantic similarity. Thus, understanding the nature of semantic similarity is central to understanding the nature of cognitive representation of lexical knowledge.

The acceptance of semantic similarity and dissimilarity is so ubiquitous that few theorists even bother to cite precedents. This has created a paradox: semantic similarity and dissimilarity are critical theoretical constructs in psychology, yet they are relatively ill-defined. The present study takes semantic similarity and dissimilarity as dependent variables. A general notion is that semantic similarity and dissimilarity of words are a critical function of the (linguistic) contexts in which words occur. This contextual approach is developed and defended in the remainder of this article.

CONTEXTUAL REPRESENTATIONS

How speakers' knowledge about word meanings is organized has attracted proposals from psychologists, linguists, and philosophers. Despite the different theoretical orientations, a common theme among them is a presumption that con-

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

texts play a critical role in the induction of the meaning of a word. Contexts can refer to nonlinguistic, situational, or other information about the communicative intents of speakers, but most studies (including this one) limit the meaning of contexts to the linguistic contexts of a word (i.e., the words preceding and following a word).

Miller and Charles (1991) advanced a contextual approach to semantic similarity that builds upon Leibniz's definition of synonymy in terms of the interchangeability of words in linguistic contexts. A main idea is that the semantic similarity of two words is a critical function of their interchangeability, without a loss of plausibility, in natural linguistic contexts. The contextual approach borrows from discussions of synonymy in terms of interchangeability. However, the contextual approach departs from the treatment of synonymy in terms of truth values, giving preference instead to plausibility. Semantic similarity and dissimilarity are gradable variables, yet there are no gradable truth values to accompany intermediate semantic similarity and dissimilarity. However, it is possible that words with different but similar meanings can be substituted for each other without creating sentences that no one is likely to use. A presumption is that the substitutability of words in contexts without a loss of plausibility is critical to their lexical similarity and dissimilarity.

Miller (1986) claimed that peoples' knowledge of a word is organized around its use in everyday discourse rather than its dictionary definition. Miller and Charles (1991) asserted that what people acquire "from hearing or seeing a word used in a series of linguistic contexts is a *contextual representation:* An association between an utterance (or inscription) and some representation of attributes common to all of these contexts" (p. 4; emphasis added). Theorists have assumed that speakers abstract some mental structure from seeing a word used in a set of linguistic contexts (Anderson & Bower, 1973). Similarly, the contextual representation of a word is not an actual linguistic context but an abstraction of information in the set of natural linguistic contexts in which a word occurs. To know the meaning of a word is to know if a given context is a member of the set satisfying the word's contextual representation.

Although the process that is used to derive a contextual representation from multiple encounters is difficult to describe, many derivatives have been characterized. Included in a word's contextual representation is information concerning its semantic, syntactic, pragmatic–stylistic, and collocational characterizations, to name a few. The semantic characterization in the contextual representation of a word is believed not to be represented as a paraphrase of a dictionary definition. Current view takes "sentences as predicate–argument construction, where verbs and adjectives serve as predicates and nouns and noun phrases serve as arguments . . . context provides information about constraints on arguments that a predicate takes, and about constraints on predicates that apply to a given argument" (Miller & Charles, 1991, p. 7). This system of constraints, mapped onto representations of the non-verbal world, yields a plausible account of one's semantic knowledge. A full account of the types of characterizations in the contextual representation of a word is found in Miller and Charles (1991).

The contextual approach is well-recognized in discussions of vocabulary development. Cruse (1986) claimed that "the meaning of a word is *constituted* by

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

its contextual relations" (emphasis added). Other linguists (Harris, 1970; Joos, 1950) have posited alternative contextual hypotheses. To assess the claim that contextual representations of semantically similar words are more similar than those of semantically different words, one needs valid and reliable data collection tools. A contextual hypothesis of meaning was first investigated in a domain of 65 synonymous nouns by Rubenstein and Goodenough (1965), using rating and co-occurrence methods of semantic and contextual similarity. In a co-occurrence test of contextual similarity, one lists all words present in the context set of item A and all words present in the context set of item B; then one computes the normalized coefficient representing the proportion of words common to both lists. The more words the lists have in common, the higher the correlation coefficient is and the more similar the two context sets are. Rubenstein and Goodenough concluded that "similarity of context is reliable as a criterion only for detecting pairs of words that are very similar . . . since words of low or medium synonymy differ relatively little in overlap" (pp. 627–633).

Miller and Charles (1991) investigated a contextual hypothesis using semantic similarity ratings and a sorting measure of contextual similarity for 6 of the 65 pairs from the Rubenstein and Goodenough study. A sorting test of contextual similarity is based on substitution (for a description, see Charles, 1988). Briefly, sets of sentences are collected for different target words and are converted to contexts by replacing targets with blank spaces. Next, a series of pairwise context sortings are given to participants, who are asked to sort the stimuli into two sets of congruous contexts (i.e., two sets of contexts capable of accepting a specific target). If two contexts are very similar, they will be difficult to tell apart. The more two sets of contexts accept either target or the more interchangeable the context sets are, the more similar they are. Sorting data are summarized by a signal detection model of contextual discriminability yielding d-primes. Lexical similarity was a critical inverse function of contextual discriminability for 5 synonyms. Thus, sorting data were more supportive of a contextual hypothesis than were co-occurrence data. In order to generalize conclusions from the sorting data, sortings of contextual similarities for the 65 pairs in Rubenstein and Goodenough were compared in terms of both semantic similarities and dissimilarities.

THE SCALING OF SEMANTIC SIMILARITY DESCRIPTORS

Typically, rating tests for eliciting semantic similarity and dissimilarity are anchored by linguistic statements that are arbitrarily chosen by the experimenters. In contrast, this study explored the feasibility of an empirical approach to choosing a series of linguistic statements (or semantic similarity descriptors) as anchors for the categories of a rating continuum.

Stimuli were a set of 14 semantic similarity descriptors that were chosen by the author based on intuitions and on usage in psychological studies of ratings for semantic similarity and dissimilarity. All stimuli appear in Appendix 1. The method of data collection was a 101-point rating continuum, ranging from 100 "identical in meaning" to 0 "opposite in meaning"; the inverse form ranged from 0 "identical in meaning" to 100 "opposite in meaning." Three groups (n =

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

12, 13, 32) of Oregon State University undergraduates were given either a random or a serial order of the 14 stimuli and were asked to write a number between 100 and 0 next to each item to indicate its meaning relative to the end points. A mean scale value per item was calculated across the different participants in a test context, yielding an average meaning for each item. The results are given in Appendix 1.

Pearson's correlation between the means of the random and the serial orders of stimuli in a direct judgment frame was $r = .947$, with slightly stronger correlations between the means of the random stimuli of the different judgment contexts ($r = -.960$) and also among the means for the serial and random stimulus orders of the different judgment contexts ($r = -.979$). All population correlations differed reliably from zero at $p < .01$.

Criteria for the selection of an explicit series of linguistic anchors for the scale categories of 5-point semantic similarity and dissimilarity ratings were developed in a two-stage process. First, the means of items for the different stimulus orders of the direct judgment context were averaged (column 5 of Appendix 1). Next, starting with the two end points, three intermediate statements were selected that were roughly equidistant (or about 25 scale units apart) from the statements directly preceding and following them on the 101-point continuum. This transformation yielded the following 5-point semantic similarity rating test, with psychometric properties analogous to an interval scale: 4 "identical in meaning" (100), 3 "similar in meaning" (75), 2 "vaguely similar in meaning" (55), 1 "different in meaning" (27), and 0 "opposite in meaning" (0). Analogous transformation of data (column 4 in Appendix 1) yielded a semantic dissimilarity rating test: 0 "identical in meaning" (0), 1 "similar in meaning" (23), 2 "distantly related in meaning" (52), 3 "dissimilar in meaning" (75), and 4 "opposite in meaning" (100). Despite different middle descriptors and different interval sizes of the semantic similarity and dissimilarity ratings, the assumptions for the selection of anchors for the rating formats were the same.

EXPERIMENT 1: RATINGS FOR SEMANTIC SIMILARITY AND DISSIMILARITY OF SYNONYMOUS NOUNS

Experiment 1 examined whether subjects' semantic similarity and dissimilarity ratings of synonymous nouns based on empirically derived scales would agree with each other and whether these ratings would agree with semantic similarity ratings of the same items based on arbitrarily derived scales used in Rubenstein and Goodenough (1965) and in Miller and Charles (1991).

Participants

Two new groups (ns = 50, 58) of Oregon State University undergraduates were studied. All received credit for courses taken in psychology. All were native speakers of English.

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

Stimuli

Stimuli were the 65 synonyms from Rubenstein and Goodenough (1965) and a subset of 30 of the 65 pairs from Miller and Charles (1991).

Method and procedure

Each set of stimuli was randomly ordered and centered on separate pages. A specific rating format was presented on a new page. The two groups of participants rated semantic similarity and dissimilarity for sets of 30 and 65 paired synonyms. Participants were advised to study each pair and to rate it for semantic similarity or dissimilarity using one of the 5-point rating formats. As an illustration of semantic similarity or dissimilarity, participants saw either a decreasing similarity or an increasing dissimilarity order of the following pairs: *snake–serpent*, *snake–eel*, *snake–alligator*, *snake–frog*, *snake–book*, and *snake–bulb*. Ratings were self-paced and took between 20 and 30 minutes.

Results and discussion

The mean semantic similarities and dissimilarities in the present study and the mean semantic similarities of the 30 and 65 synonyms from the arbitrary ratings formats in Miller and Charles (1991) and Rubenstein and Goodenough (1965), respectively, were in good correspondence. However, the inverse correlation between the semantic similarities and dissimilarities of the 30 synonyms in the present study ($r = -.967$) slightly exceeded a direct correlation between the sets of semantic similarities of the 30 synonyms based on the ratings in the earlier studies ($r = .956$). In contrast, the direct correlations between the sets of semantic similarities of the 30 synonyms from the present study's ratings and from the ratings in Miller and Charles and in Rubenstein and Goodenough declined to $r = .938$ and $r = .929$, respectively. The inverse correlations between the rated semantic dissimilarities in the present study and the rated semantic similarities in Rubenstein and Goodenough and in Miller and Charles declined to $r = -.919$ and $r = -.916$, respectively. All population correlations differed reliably from zero at $p < .01$.

In addition, the size of the correlation between rated semantic similarities and dissimilarities of the present study reliably exceeded the size of the correlations between rated semantic dissimilarity of the present study and each set of rated semantic similarity of the earlier studies ($t = 2.34$, $df = 37$, $p < .05$). It did not differ significantly from the size of the correlations between the semantic similarity ratings of the present study and those of the earlier studies ($t = 1.73$, $df = 27$, $p < .05$).

Generally, the intracorrelations between semantic similarities and dissimilarities from empirically anchored ratings were stronger than the intracorrelations between semantic similarities from arbitrarily anchored ratings. They were also stronger than the intercorrelations between the relatedness data for empirically and arbitrarily anchored ratings. In some cases, the difference was reliable.

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

EXPERIMENT 2: SORTING SENTENTIAL CONTEXTS OF SYNONYMOUS NOUNS

Experiment 2 employed a sorting test to elicit contextual similarities of the 65 synonymous nouns. The main purpose of the study was to determine more completely the relation between substitutability and similarity of meaning for the 65 synonyms and thus to extend the sorting data for the subset of 6 synonymous nouns in Miller and Charles (1991). Evidence for the reliability of data from the current pen-and-paper format of the present study and the card-sorting format of Miller and Charles was explored.

Participants

Participants were 115 Oregon State University undergraduates who had not taken part in the rating test construction or in the relatedness ratings in Experiment 1. All were native speakers of English and obtained credit for courses taken in psychology.

Materials

Sets of subject-generated contexts were elicited for each of the 48 target nouns constituting the 65 synonyms, in the tradition of Rubenstein and Goodenough (1965) and Miller and Charles (1991). Each of the target nouns was written on a separate page. Different sets of 9 or more participants wrote three to four sentences, ten words or longer, for each of the target nouns. This created a corpus of 25 to 40 sentences for each noun. A representative set of 15 sentences for each target was then converted to contexts by replacing the target with an underline. By eliciting the contexts of each target word from different participants, the possibility of spurious overlap was minimized. Illustrative sentence contexts for some concepts include the following.

1. One sign of being famous or popular is when people come over and ask you for your_____.(autograph)
2. Mechanized productions and unions have grown a great deal in the past 75 years, largely due to the_____.(automobile)
3. The_____was perched on a long black telephone wire singing very loudly. (bird)
4. I did not wash my_____today because it was raining in Corvallis.(car)
5. When I was a child the most scary place I could think of was the_____ because people were buried there.(cemetery)
6. Apples and bananas are two_____that can't be compared with each other. (fruits)
7. Take this paper home and get a_____from your parents allowing this field trip.(signature)

In Miller and Charles (1991), subject-generated sentence contexts optimized the elicitations of the semantic space of the 6 synonym nouns, unlike the senten-

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

tial contexts in the Brown Corpus (Francis & Kucera, 1982). Subject-generated sentence contexts were also found to optimize elicitations of the prototype structure of emotion words, compared to sentential contexts in introductory psychology texts (Fehr & Russell, 1984).

Procedure

The 30 sentence contexts for each of the 65 paired sorting tests were haphazardly ordered by the experimenter using a computer. Each pairwise sorting test was printed on a separate page; two targets were written on top of each page. Participants were asked to substitute one of the two target words in the place of the underline in each context to produce a sensible sentence. Responses were written at the end of each sentence. Participants examined a series of four to five paired context sortings for concepts of varying synonymity, moving on to a new pair after finishing each sort. A set of contexts was seen once by the participants, and between 9 and 24 participants took part in the different pairwise sortings.

Results

Appendix 2 illustrates sortings for 15 of the 65 synonymous nouns in Experiment 1. The sorting data are displayed in order of increasing semantic dissimilarity. As there were 15 contexts per target and 9 to 24 subjects in the various sortings, 135 to 360 context sortings were possible (15 contexts $\times$ 9–24 subjects = 135–360). Total context sortings in the different paired comparisons were between 270 and 720.

Consider the contexts of the concepts *gem–jewel*, a pair typically rated among the least dissimilar or most similar of the stimuli. The 15 *jewel* contexts were sorted a total of 180 times by 12 participants: 84 (47%) correct sortings of *jewel* and 96 (53%) mistaken elicitations of *gem*. The 15 *gem* contexts produced 180 total sortings from the same 12 participants: 112 (62%) correct sortings of *gem* and 68 (38%) mistaken elicitations of *jewel*. Response bias for *gem* was modest. The other sortings were understood in the same way.

In general, sortings of contexts of the least dissimilar or most similar concepts were the most interchangeable, whereas interchangeability declined for moderately dissimilar concepts (*coast–forest*). It declined to baseline levels for very dissimilar concepts (*noon–string*). Thus, the relation between substitutability and meaning for the 30 and 65 synonymous nouns was straightforward. The substitutability of contexts was indirectly related to semantic dissimilarity or, alternatively, the substitutability of contexts and semantic similarity varied directly.

Response bias in the sortings yielded evidence of the dominance relations among synonyms. As in Miller and Charles (1991), a small response bias favored the hyponym *fruit* over the generic term *food*. However, the generic term *bird* was strongly favored over the hyponym *cock*. Charles (1988) noted a strong response bias for the generic term *building* over the hyponym *hotel*. Thus, preference for generic terms over hyponyms in the sortings was not a consistent

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

feature of pairs related by hyponymy. Response bias for the 6 concepts common to Miller and Charles and the present study was similar. Response bias in the sortings was attributed to word frequency (Charles, 1988).

In Appendix 2, the contexts of *bird* and *crane* were less often interchanged than were those of *bird* and *cock*, despite the closer ratings of the former than the latter. *Webster's New Riverside University Dictionary* indicates that *crane* and *cock* are polysemous, with an animate sense and an inanimate sense. In this study, the *crane* context ("The_____has the capability of lifting tons of weight") expressed an inanimate sense, but the *cock* context ("A farm in Oregon breeds _____for fighting and was discovered by the police") and the *bird* context ("The_____was perched on a long, black telephone wire singing very loudly") expressed similar animate senses. Thus, the *crane–bird* contexts were more distinguishable than the *cock–bird* contexts. Paired ratings of *bird–cock* and *bird–crane* presumably activated the animate sense of *crane* and *cock* – hence the close overlap. However, isolated presentations of *cock* and *crane* in the generation of contexts activated their animate and inanimate senses, respectively (presumably the more familiar or frequent senses). Such results confirm the assumption in Charles (1988) that linguistic contexts can disambiguate the meanings of polysemous words. The sorting test also appears valid for assessing the extent to which sense relations between lexical concepts are reflected in their constituent contexts.

Signal detection model of contextual discriminability

Sorting data are conveniently summarized by a signal detection model (Egan, 1975) of contextual discriminability, yielding d-primes. For example, consider the pair *gem–jewel*. If *gem* is taken as the signal to be detected, the probability of a hit – P("gem"|*gem*) (i.e., the probability of sorting *gem* context under the label "gem") – is estimated as 84/180 or 0.467%; the probability of a false alarm – P("gem"|*jewel*) (i.e., the probability of sorting *jewel* context under the label "gem") – is estimated as 68/180 or 0.378%. These probabilities correspond to a d-prime of 0.40. Alternate calculations for *jewel* as the signal yield essentially the same d-prime. Similar calculations of d-primes were made for all 65 pairwise sortings; these are illustrated for the subset of 30 concepts in Table 1.

Discussion

Pearson's correlations between lexical similarity and dissimilarity based on empirically and arbitrarily anchored ratings and also between the d-primes and various rating data appear in Table 2. The data in Table 2 reinforce the conclusions drawn from an inspection of the raw data in Appendix 2. A linear direct correlation between the d-primes and the semantic dissimilarities of the 65 pairs in the present study was strong. It was replicated by a strong linear inverse correlation between the d-primes and the semantic similarities of the 30 synonyms. The results generalized to linear inverse correlations between the d-primes and the semantic similarities for the sets of 30 and 65 synonyms, based on arbitrarily anchored ratings of the earlier studies. All population correlations

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

Table 1. *Ratings for semantic dissimilarity and similarity and sorting of d-primes for the 30 synonymous nouns*

Synonymy pair	Dissimilarity	Similarity	Rubenstein and Goodenough (1965)	Miller and Charles (1991)	d-primes
Midday–noon	0.40	4.00	3.94	3.42	1.35
Automobile–car	0.44	3.89	3.92	3.92	0.85
Gem–jewel	0.56	4.00	3.94	3.84	0.40
Boy–lad	0.56	3.62	3.82	3.76	1.17
Coast–shore	0.66	3.24	3.60	3.70	1.57
Journey–voyage	0.70	3.64	3.58	3.84	0.75
Magician–wizard	0.80	3.38	3.21	3.50	1.93
Furnace–stove	0.94	2.97	3.11	3.11	2.40
Bird–crane	1.20	2.91	2.63	2.97	4.30
Food–fruit	1.34	2.74	2.69	3.08	2.73
Bird–cock	1.46	2.72	2.63	3.05	1.62
Asylum–madhouse	2.00	3.10	3.04	3.61	3.42
Brother–lad	2.00	2.17	2.41	1.66	>4.10
Tool–implement	2.02	2.48	3.66	2.95	1.20
Brother–monk	2.04	2.02	2.74	2.82	4.75
Coast–forest	2.30	2.00	0.85	0.42	3.10
Car–journey	2.34	1.70	1.55	1.16	3.90
Shore–woodland	2.48	1.09	0.90	0.63	3.68
Monk–oracle	2.64	1.58	0.91	1.10	2.35
Food–rooster	2.70	1.48	1.09	0.89	5.12
Coast–hill	2.80	1.14	1.26	0.87	3.95
Crane–implement	2.82	1.31	2.37	1.68	4.43
Cemetery–woodland	2.90	1.36	1.18	0.95	3.33
Monk–slave	2.92	1.26	0.57	0.55	3.70
Forest–graveyard	2.94	1.24	1.00	0.84	3.57
Lad–wizard	2.94	1.07	0.99	0.42	4.02
Cord–smile	2.98	1.03	0.02	0.13	>4.10
Glass–magician	3.04	1.05	0.44	0.11	4.95
Rooster–voyage	3.06	0.33	0.04	0.08	4.75
Noon–string	3.08	0.95	0.04	0.08	>5.20

Note: Mean ratings on a scale of 0 to 4 for dissimilarities and similarities and sorting of d-primes from 0 to >5.55 for sets of 30 of 65 synonyms (Charles, 1988; Miller & Charles, 1991; Rubenstein & Goodenough, 1965). Pairs are ordered by increasing dissimilarity.

were statistically different from zero at $p < .01$. The relations between the d-primes and the semantic similarities and dissimilarities of the 65 synonyms from the earlier studies are graphed in Figures 1 and 2.

Such reliable correlations between the d-primes and the semantic similarities and dissimilarities of the sets of synonymous nouns, independent of rating format type, are consistent with the assumptions of a contextual hypothesis of meaning. That is, semantic similarity and dissimilarity are critical correlates of

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

Table 2. *Matrix of correlations between semantic similarity and dissimilarity from empirical and arbitrary rating methods and d-primes for contextual similarity from a sorting test*

Test type	(1)	(2)	(3)	(4)	(5)
(1) Rubenstein & Goodenough	—	.956	.929	–.919	–.822
(2) Miller & Charles		—	.938	–.916	–.720
(3) Similarity			—	–.967	–8.18
(4) Dissimilarity				—	8.26
(5) d-primes					—

Note: Rubenstein and Goodenough (1965) and Miller and Charles (1991) are tests of semantic similarity based on arbitrarily anchored ratings for 65 and 30 synonyms, respectively. Similarity and dissimilarity are from current rating formats with empirically chosen anchors. The d-primes are from a sorting test of contextual similarity. All correlations were reliable at $p < .01$.

contextual similarity. The data replicated and generalized the context sortings for the 6 synonyms studied in Miller and Charles (1991). Thus, confidence in the validity and generality of the assumptions of a contextual hypothesis is strengthened.

This measure of contextual similarity based on substitutability provides a better estimate of semantic similarity than does a measure of contextual similarity based on co-occurrence (Rubenstein & Goodenough, 1965). In Figures 1 and 2, the relation between meaning and substitutability is remarkably close, despite the different rating data from the present and earlier studies – a fact borne out by the close correlations in Table 2. The sluggish rate of change between the middle and far similarities and co-occurrences found in Rubenstein and Goodenough is replaced by a more dynamic rate of change or a strong linear relation between meanings and d-primes across the entire continuum. Indeed, contextual similarities appear to change more slowly at the upper meanings than at the middle or lower meanings. Thus, the co-occurrence measure of contextual similarity in Rubenstein and Goodenough mainly contributes to the limited support for a contextual hypothesis.

In Figures 1 and 2, the scatter or deviation from linearity is mostly at the middle levels of meanings, which prompts some interesting speculations. For example, three deviant cases in the upper left quadrant of Figure 1 (*oracle–sage*, *bird–cock*, and *sage–wizard*) involved very low frequency or polysemous concepts. Similarly, *magician–oracle* and *monk–oracle* in the lower left quadrant of Figure 1 were deviant cases from the normally linear relation. *Brother–monk* and *bird–crane* in the upper right quadrant of Figure 1 were also among the more deviant cases. These terms as well as others were deviations from linearity in Figure 2: *bird–crane*, at the extreme lower right quadrant, and *oracle–sage* and *tool–implement*, at the extreme upper left quadrant. Such exceptions involved polysemous or low frequency concepts that were often rated moderately similar or dissimilar, while their d-primes typically were either smaller or larger than the ratings suggested. It would appear that either a lack of explicit informa-

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

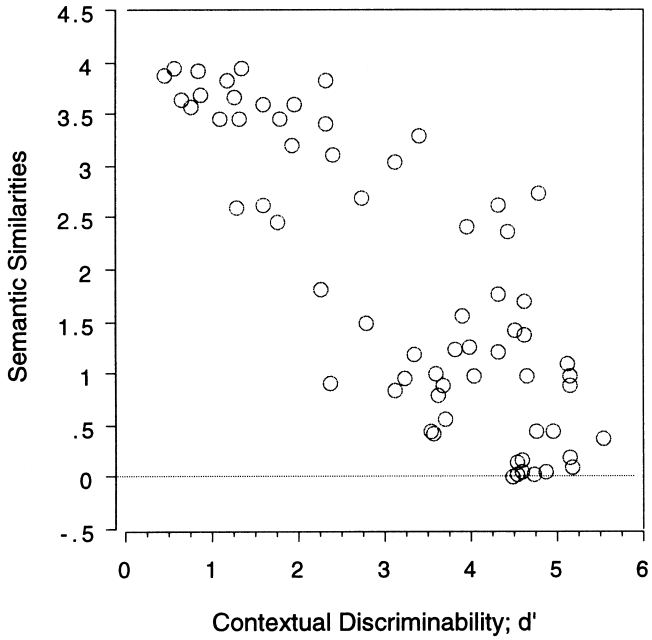


Figure 1. Semantic similarities as a function of contextual discriminability of the 65 synonyms.

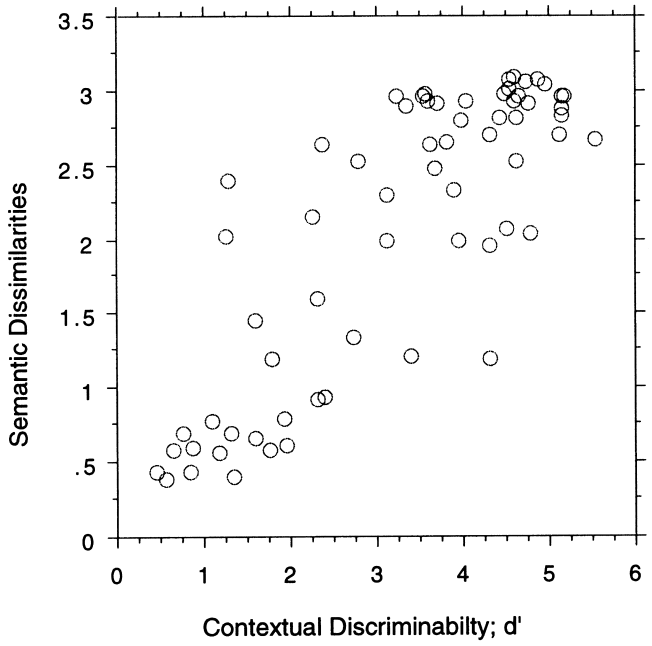


Figure 2. Semantic dissimilarities as a function of contextual discriminability of the 65 synonyms.

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

tion about which sense of a concept is relevant or an unfamiliarity with the concepts due to their low frequency led participants to adopt a conservative strategy, rating such concepts as neither closely nor remotely similar or dissimilar. If this is so, a test based on linguistic contexts would provide a greater resolution of meanings and access to semantic memory than one based on ratings of lexical concepts.

The data supported the conclusion that different sorting formats based on substitution elicit valid and reliable estimates of contextual similarity. The positive correlations between the estimates of contextual similarity for 6 of the 65 nouns from the present pen-and-paper format and the prior card-sorting format were strong. They ranged from a high of $r = .978$ between the two card-sorting sessions in Miller and Charles (1991) to $r = .974$ and $r = .950$ between the present d-primes for the pen-and-paper format and the prior d-primes for the card-sorting format. All correlations between the d-primes of synonymous nouns from the sorting formats differed reliably from zero at $p < .01$.

GENERAL DISCUSSION

A contextual hypothesis of meaning presupposes that conceptual similarity is a critical correlate of contextual similarity. These reliable correlations between the d-primes and the semantic similarities and dissimilarities of 30 and 65 synonymous nouns, independent of empirically and arbitrarily anchored rating tests, considerably strengthened the argument for a contextual hypothesis. The data also replicated and generalized sorting data and supported arguments for a contextual hypothesis based on the findings of Miller and Charles (1991). Evidence of the reliability and validity of new rating and sorting methods of eliciting lexical and contextual similarity was found.

Rating and sorting methods of data collection

Likert-type rating methods (Likert, 1932) are a ubiquitous data collection tool in the psychological sciences for ordering the intensity of gradable variables. In contrast to arbitrarily anchored ratings, empirically anchored ratings of semantic similarity and dissimilarity are more consistent with the spirit of the original Likert scale. An empirical approach to the construction of semantic relatedness rating tests generally optimized efficiency in participants' responses, unlike the arbitrary rating formats. Compared to arbitrarily anchored rating formats, the scaling of semantic similarity descriptors yielded verbal anchors for rating tests for eliciting semantic similarity and dissimilarity that were more consistent with speakers' approach to the task. Even greater dissociation in the size of correlations of semantic similarities and dissimilarities, based on empirical and arbitrary rating formats, was found in the semantic field of emotions (Charles, 2000). Clearly, empirically anchored rating methods for eliciting semantic similarity and dissimilarity deserve further consideration.

A sorting test of contextual similarity also must be considered a valid alternative to a traditional co-occurrence method. Psychologists are probably familiar with this co-occurrence technique from Deese's (1964) use of "intersection coef-

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

ficients" to calculate the overlap of two distributions of word associations. Intersection coefficients for word associations generally support the assumption that shared associates have similar meanings. Paradoxically, Deese found the strongest intersection coefficients for word associations of antonymous adjectives. This led to a counterintuitive conclusion that antonymy is a type of semantic generalization. In contrast, the d-primes for context sortings of antonymous adjectives (*strong–weak*) were progressively larger than the d-primes for contexts sortings of indirect antonyms (*strong–faint*) and similar adjectives (*strong–powerful*). Thus, sorting data led to the intuitive conclusion that direct antonyms are less interchangeable (or less similar) in natural contexts than are synonymous adjectives and nouns. A major disadvantage of a substitution test, unlike a co-occurrence test, is that there is no quick and easy algorithm for its calculation, and it is cumbersome to administer. However, a pen-and-paper format minimizes the inefficiencies in test administration and scoring evident in the earlier card-sorting formats.

Contextual correlates of meaning

These data strongly favor the claim that judgments of semantic similarity and dissimilarity of words are critically related to the similarity of their contextual representations. Following Miller and Charles (1991), the results suggest that to know the meaning of a word is to know how to use it, and that encounters with the natural linguistic contexts of a word yield a contextual representation that summarizes information about the contexts in which the word may be used. The notion of a contextual representation of a word borrows from a common argument in linguistics and in psychology that most word meanings are derived from (linguistic) contexts. It is also consistent with the so-called schematic or prototype models of word meaning and with top-down processing models in reading comprehension and discourse processing. The important point is that people consolidate the salient information from a word's contexts into a single representation that summarizes them.

Sorting data for nouns and adjectives (Charles & Miller, 1989) point to a dissociation in the cognitive representations of nouns and adjectives. Although hyponymy between noun concepts is critical, associative bipolarity of antonymous adjectives has been emphasized in some models (Charles & Miller, 1989; Charles, Reed, & Derryberry, 1994). Psychologists such as Murphy and Andrew (1993) have favored a conceptual approach to antonymy and synonymy. Deese's (1964) finding of strong intersection coefficients for the word association distributions of antonymous adjectives has fueled arguments that antonyms are not as different in meaning as they seem. As word association distributions supposedly represent conceptual similarity, it is assumed that the underlying representations for antonyms are conceptual in nature. Such associations of antonyms are also believed to result from the substitutability of antonyms in linguistic contexts (Ervin, 1961; Juteson & Katz, 1989).

The asymmetrical substitutability of contexts of direct synonyms in the present study and of synonymous adjectives in Miller and Charles (1989) as well as contexts of adjectival direct antonyms in Charles and Miller (1989) and of

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

antonymous noun emotions in Charles (2000) represent a contradiction to conceptual arguments of antonymy. The present author takes this to mean that the cognitive representation of antonyms is associative bipolar and that of synonyms is conceptual in nature. In support of this argument, Fellbaum (1995), Miller and Charles (1989), and Charles (2000) showed that associative antonyms cooccur in linguistic contexts to a greater degree than chance predicts. Given this fact, the patterns of word associations between antonymous adjectives deserve reinterpretation.

Alternative support for a contextual approach to meaning comes from the computational model of meaning known as latent semantic analysis (LSA) (Landauer, 1998; Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). LSA is a theory and method for extracting and representing the contextualized meanings of words by statistical computations of large corpora of text. The main idea is that the aggregate of all the word contexts in which a word can and cannot occur provides a set of constraints that determines the similarity of meaning of words. Cell entries in a matrix of rows and columns symbolize the frequencies with which a word type of a row appears in texts or contexts denoted by a column. LSA uses singular value decomposition to infer much deeper and emergent semantic representations from a text than is possible with typical cooccurrence models. The method of singular value decomposition also may be useful for understanding the process by which speakers abstract a contextual representation from multiple encounters with a word's linguistic contexts. The psychological reality of LSA has been demonstrated. Its scores overlap with those of humans on vocabulary and subject matter tests, and it mimics human word sortings, category judgments, and lexical priming. However, LSA fails to utilize perceptual data, feelings, and intentions, and it does not incorporate word order or syntax and morphology (although the HAL and KDC incorporate word order).

The success of the experimental and computational models of contextualized meanings without any explicit recourse to primitive features is noteworthy. It is possible that similar contextual representations underlie judgments of similarity-based phenomena as iconic, analogical, and connotative. Charles (2000) explored the intersection between contextual representations and the connotative similarity of antonymy and synonymy of emotions. It would be useful to compare computational and experimental approaches to contextualized meanings in a semantic field and across several semantic relations.

CONCLUSIONS

Reliable correlations between semantic similarity and dissimilarity and sortings of contextual similarity of a set of 65 synonymous nouns considerably strengthened the claim that the more similar two words are, the more similar their linguistic contexts are. The results also suggested that, as speakers learn the meaning of a word from encountering its natural linguistic contexts, they acquire a schematic contextual representation of the linguistic contexts of a word. Speakers invoke a contextual representation of a word to determine the plausibility of substituting a word into a specific context and to determine the per-

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

ceived lexical similarity and dissimilarity of synonymous nouns. The data also converged with data from a computational model of contextualized meaning. It is quite possible that similar contextual representations underlie the perceptions of other types of similarity and semantic relations. Present methods of eliciting lexical and contextual similarity of concepts appear superior to alternative methods of data collection generally in use.

APPENDIX 1

Mean scale values for semantic similarity descriptors from two judgment modes

	Direct		Inverse	
Semantic similarity descriptors	Random	Serial	Random	Pooled
Identical in meaning	100.00	100.00	0.00	100.00
Strongly similar in meaning	91.90	90.40	17.58	91.15
Strongly related in meaning	88.10	86.25	10.84	87.18
Very similar in meaning	84.50	86.93	12.00	85.72
Similar in meaning	77.41	72.50	23.08	74.96
Very much alike in meaning	71.25	83.33	15.84	77.29
Vaguely similar in meaning	58.00	51.76	40.00	54.88
Distantly related in meaning	52.00	41.92	52.25	46.96
Unrelated in meaning	44.30	19.17	77.50	31.74
Dissimilar in meaning	37.50	35.83	75.50	36.67
Having nothing to do with each other	23.90	10.17	79.00	17.04
Different in meaning	20.40	33.08	72.16	26.75
Very dissimilar in meaning	18.25	18.33	80.90	18.24
Highly dissimilar in meaning	13.08	20.83	86.50	16.96
Somewhat opposite in meaning	2.25	5.77	92.84	4.01
Opposite in meaning	0.00	0.00	100.00	0.00

Note: Means for 12 and 13 subjects in different orders of a direct judgment frame and for 32 subjects in an inverse judgment frame. The means of scores for random and serial orders of direct judgment were pooled. Items are ordered by the pooled data.

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

APPENDIX 2

Illustrative sorting data for contexts of synonymous nouns

Stimulus	Responses		Sum
	Jewel	Gem	
Jewel	84	96	180
Gem	68	112	180
Sum	152	208	360
	Midday	Noon	
Midday	96	84	180
Noon	19	161	180
Sum	115	245	360
	Auto	Car	
Automobile	94	101	195
Car	42	153	195
Sum	136	254	390
	Coast	Shore	
Coast	135	45	180
Shore	33	147	180
Sum	168	192	360
	Journey	Voyage	
Journey	136	59	195
Voyage	80	115	195
Sum	216	174	390
	Bird	Crane	
Bird	187	8	195
Crane	1	194	195
Sum	188	202	390
	Food	Fruit	
Food	187	23	210
Fruit	14	196	210
Sum	201	219	420
	Bird	Cock	
Bird	283	32	315
Cock	115	200	315
Sum	398	232	630

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

APPENDIX (*continued*)

Stimulus	Responses		Sum
	Coast	Forest	
Coast	244	26	270
Forest	10	260	270
Sum	254	286	540
	Car	Journey	
Car	297	3	300
Journey	18	282	300
Sum	315	285	600
	Food	Rooster	
Food	194	1	195
Rooster	1	194	195
Sum	195	195	390
	Monk	Slave	
Monk	192	3	195
Slave	12	183	195
Sum	204	186	390
	Smile	Chord	
Smile	204	6	210
Cord	0	210	210
Sum	204	216	420
	Glass	Magician	
Glass	299	1	300
Magician	4	296	300
Sum	303	297	600
	Noon	String	
Noon	195	0	195
String	0	195	195
Sum	195	195	390

Note: Cells contain the number of times the contexts of a word type were sorted correctly and incorrectly for 15 of the 65 synonyms (Rubenstein & Goodenough, 1965). Cells are ordered by increasing dissimilarity.

ACKNOWLEDGMENTS

This research was supported by NIH MBRS Grant 2 S06 GM 08182-16A to Delaware State University. The data were collected when the author was an assistant professor at Oregon State University, Corvallis. The author thanks Michael Sharon, an undergraduate at Oregon State University, for assistance in data collection. The author also thanks several anonymous reviewers for invaluable advice.

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

REFERENCES

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Winston.
Carey, S. (1985). *Conceptual changes in childhood*. Cambridge, MA: MIT Press.
Charles, W. G. (1988). The categorization of sentential contexts. *Journal of Psycholinguistic Research, 17*, 401–403.
(2000). *Contexts of antonymous and synonymous emotion nouns*. Manuscript in preparation.
Charles, W. G., & Miller, G. A. (1989). Contexts of antonymous adjectives. *Applied Psycholinguistics, 10*, 357–375.
Charles, W., Reed, M. A., & Derryberry, D. (1994). Conceptual and associative processing of antonymy and synonymy. *Applied Psycholinguistics, 15*, 331–356.
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.
Egan, J. P. (1975). *Signal detection theory and ROC analyses*. London: Academic.
Ervin, S. M. (1961). Changes with age in verbal determinants of word-association. *American Journal of Psychology, 74*, 361–367.
Fehr, B., & Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General, 113*, 464–486.
Fellbaum, C. (1995). Co-occurrence and antonymy. *International Journal of Lexicography, 8*, 281–303.
Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mufflin.
Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 225–277). Cambridge: Cambridge University Press.
Glick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.
Gross, D., Fischer, U., & Miller, G. A. (1989). The organization of adjectival meanings. *Journal of Memory and Language, 28*, 92–106.
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.
Joos, M. (1950). Description of language design. *Journal of the Acoustical Society of America, 22*, 701–708.
Juteson, J., & Katz, S. (1989). *Co-occurrences of antonymous adjectives and their contexts*. Yorktown Heights, NY: IBM Research Report.
Landauer, T. K. (1998). Learning and representing verbal meaning: The latent semantic analysis theory. *Current Directions in Psychological Science, 7* (8), 161–164.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problems: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*, 211–240.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25*, 259–284.
Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*, 5–55.
Miller, G. A. (1986). Dictionaries in the mind. *Language and Cognitive Processes, 1*, 171–185.
Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes, 6*, 1–28.
Murphy, G. L., & Andrew, M. J. (1993). The conceptual basis of antonymy and synonymy in adjectives. *Journal of Memory and Language, 32*, 301–319.
Osgood, C. E., Suci, J. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
Quine, W. V. O. (1953). Two dogmas of empiricism. Reprinted in W. V. O. Quine, *From a logical point of view* (pp. 20–46). Cambridge, MA: Harvard University Press.
Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic.
Rubenstein, H., & Goodenough, J. B. (1965). Contextual correlates of synonymy. *Computational Linguistics, 8*, 627–633.

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press

Smith, E. E., Shobin, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*, 214–241.
Tulving, E., & Thompson, S. (1975). Encoding specificity and retrieval process in episodic memory. *Psychological Review, 80*, 352–373.
Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

https://doi.org/10.1017/S0142716400004057 Published online by Cambridge University Press