



*George A. Miller*

# ON KNOWING A WORD

*George A. Miller*

Psychology Department, Princeton University, Princeton, New Jersey 08544;  
e-mail: geo@clarity.princeton.edu

KEY WORDS: context, lexical semantics, polysemy, sense identification, WordNet

---

## ABSTRACT

A person who knows a word knows much more than its meaning and pronunciation. The contexts in which a word can be used to express a particular meaning are a critical component of word knowledge. The ability to exploit context in order to determine meaning and resolve potential ambiguities is not a uniquely linguistic ability, but it is dramatically illustrated in the ease with which native speakers are able to identify the intended meanings of common polysemous words.

---

## CONTENTS

INTRODUCTION .....	1
CRITERIA FOR WORD KNOWLEDGE .....	2
SEMANTICS .....	3
<i>Lexical Semantics</i> .....	5
<i>WordNet</i> .....	8
CONTEXTUALIZATION .....	10
<i>Polysemy</i> .....	11
<i>Contextual Representations</i> .....	13
<i>Automatic Word-Sense Identification</i> .....	15
CONCLUSION .....	17

## INTRODUCTION

The question of what it means to know a word has fascinated many psychologists, sometimes with consequences of considerable practical value to the science. Among historically more important examples one thinks of Galton (1879) experimenting with word associations, Binet (1911) using word tasks

to test mental age, Thorndike (1921) trying to determine what words every high school graduate should know, Ogden (1934) selecting 850 words to define Basic English, or Bühler (1934) analyzing deictic words that acquire referential value from the contexts in which they are used. Clearly, word knowledge has provided fertile ground for psychological studies; the reason may be that it raises fundamental questions. That is to say, knowing a word is generally considered to be a matter of knowing the word's meaning, and meaning is one of those concepts of great importance for understanding the nature and limits of psychology.

My own interest in word knowledge grew out of an abiding curiosity about communication in general and linguistic communication in particular. Initially, my interest in words was part of a dislike for the behavioristic emphasis on their ostensive or referential function, an emphasis that seemed to ignore their relational or inferential aspects. It seemed obvious to me that words are interrelated in diverse and complex ways, and I explored several lines of evidence that testified to the psychological reality of those lexical interrelations.

As I have learned more about the lexical component of language, however, my reasons for continuing to study it have changed. Today I am fascinated by the human ability to contextualize, the ability to use context to determine meanings and to resolve potential ambiguities (Miller 1996). Whenever experience is meaningful, contextualizing is involved. People object to being quoted out of context, but nothing can be out of context; the objection is to being quoted in the wrong context. Contextualization is not exclusively linguistic, of course; using context to determine linguistic meaning is simply a special case of a general cognitive ability. But the easiest place to study contextualization is surely in the relation between words and their contexts of use.

Knowing a word involves knowing its meaning and therefore, in my view, knowing a word involves knowing its contexts of use. So my present concern is how to characterize that contextual knowledge; this chapter develops the question and reports how far I have come.

## CRITERIA FOR WORD KNOWLEDGE

What is involved in knowing a word? One interpretation of this question is performative. It becomes: What tasks does knowing a word enable a person to perform? The answer gives a set of tasks that can provide evidence that a person really knows a word. This interpretation of the question has received much attention.

Psychologists have used various methods to determine whether a person knows a word (Beck & McKeown 1991). The simplest is to ask yes-or-no questions: "Do you know the meaning of  $x$ ?" (Seashore & Eckerson 1940). Binet & Simon (1911) asked younger children to name objects or describe pic-

tures; older children were asked to give definitions or to use the words in sentences. It is not clear who should be credited with inventing multiple-choice questions, but recognition of an associated phrase has become the most frequently used method to test word knowledge. And since none of these methods is infallible, they are sometimes used in combination. For example, Anglin (1993) first asked children for a definition; failing that, he asked for a sentence using the word, plus an explanation of what the word meant in that sentence; failing that, he used a multiple-choice question. Subjects were given credit for knowing the word if they succeeded at any of these tasks.

There are, of course, different levels of word knowledge between complete innocence and complete competence. The recognition vocabulary is much larger than the speaking vocabulary, for example. Cronbach (1942) described different kinds of knowledge of a word as the ability to define it, the ability to recognize situations for using it, knowledge of its alternative meanings, the ability to recognize inappropriate uses of the word, and the availability of the word for use in everyday life.

All of which is both true and useful. Yet, leaving aside various marginal states of knowledge, to say that a person who knows a word can use it correctly does not really describe what is involved in knowing a word. It is hard to avoid a feeling that problem and method have passed one another by.

What is involved in knowing a word? Another interpretation of that question is substantive. Then it becomes: What does a person who knows a word know? A ready answer is that a person who knows a word must know its meaning(s).

## SEMANTICS

The study of meanings is semantics. The kinds of semantics are so diverse, however, that some rough distinctions are needed to render the topic manageable.

One important distinction is between logical and linguistic semantics. Logical semantics attempts to formulate a general theory of the set of meanings that a language can express, an axiomatized theory (usually for some logical or mathematical language) that can be shown to be consistent and complete. The semantics of Montague (1974) is a notable example of the application of those methods to a natural language. Linguistic semantics, on the other hand, is descriptive, a characterization of the meanings that have been expressed in natural languages; the semantics of Jackendoff (1972) and of Lyons (1977) are notable examples, although Lyons borrows more from logical semantics than Jackendoff does.

A second important distinction is between sentential and lexical semantics. Sentential semantics is concerned with the meaning of statements; because

first-order logic can be applied to statements, this approach is sometimes favored by logicians. An example would be a verification theory of semantics. A verification theory might assume that the meaning of any sentence is the set of individually necessary and jointly sufficient conditions for the sentence to be true; then it might be further assumed that anyone who knew those conditions would understand the meaning of the sentence. Considering that isolated words are not generally true or false, they are, as far as sentential semantics is concerned, meaningless. Lexical semantics, on the other hand, assumes that words in isolation do have meaning; a sentence acquires its meaning by virtue of the words that compose it and the manner of their combination. This approach is generally favored by linguists; the lexical semantics of Cruse (1986) is a notable example.

The assumption that words have meanings presupposes an assumption, familiar since de Saussure's (1916) distinction between *signifie* and *signifiant*, that words are doubly entered in lexical memory, once phonologically and once semantically, with associations between them. Both the phonological (or orthographic) form and the concept that the form can be used to express must be learned. The word form is learned by hearing (or reading) it and by uttering (or writing) it; the word meaning may be ostensively grounded in immediate perceptual experience (as the concepts expressed by such words as *dog*, *green*, *fire*, *moon*, *rain*, and the like), or in some more complex manner (as the concepts expressed by such words as *democracy*, *mendacity*, *God*, *psychology*, or *climate*); whether all abstract words can be reduced to (or "grounded in") empirical experience is a question that long books have been written about (Miller & Johnson-Laird 1976). But it is generally assumed that words that can be used referentially (nouns, verbs, adjectives, adverbs) acquire their referential value by virtue of the concepts that they can express, and that this referential value is an important part of a language user's lexical knowledge.

Word form and word meaning are normally learned together. Indeed, one might define a word as an association between a word meaning and a word form that plays a syntactic role. The lexicon, on this view, is simply a list of all the form-meaning associations in the language—what Pustejovsky (1995) has called a "sense enumeration lexicon." When a word form is associated with more than one word meaning, it is said to be polysemous; when a word meaning can be expressed by more than one word form, those word forms are said to be synonymous.

The assumption that knowing a word entails knowing what the word means places the present topic squarely in the area of lexical semantics. But a caveat: To assume that word forms are associated with word meanings does not entail that everyone who uses the language will be able to explain those meanings. People communicate via sentences, seldom via isolated words. Consequently, people's understanding of the meaning of sentences is far more reliable than

their understanding of the meaning of words. Their intuitions about the definitions of the words they utter and understand are fragmentary at best. Lexical semanticists try to tap semantic intuitions of native speakers by embedding questions about the meaning of a word in sentential contexts. But it is delicate work, often involving the recognition not of a word's meaning, but of a semantic anomaly. What people know when they know the meaning of a word is more a skill than a fact, the skill of incorporating that word appropriately into meaningful linguistic contexts.

Finally, a distinction should be noted between lexical semantics and lexicography. It is a lexicographer's task to compile associations between word forms and word meanings and to publish them in a convenient alphabetical list. Although lexical theory and lexicographic practice are ordinarily close, in addition to deciding what should be included in a lexical entry and how definitions should be written, lexicographers must actually write the definitions and at the same time worry about such practical matters as how to keep their dictionary down to a manageable size (Landau 1984). Commercial considerations aside, however, most lexicographers would argue that a good dictionary is the ultimate test of any theory of lexical semantics; they sometimes become impatient with criticisms from theorists who have never worked as lexicographers. As one lexicographer has put it, "most of the words one has to deal with when working through the alphabet turn out to be more recalcitrant than those chosen as examples in works on semantic theory" (Atkins 1992/1993:19).

Nevertheless, there is a consensus that the sense enumeration found in a standard dictionary is not the kind of lexicon required for linguistic or psycholinguistic theory. In particular, the domains that words refer to are themselves highly structured in ways that a simple alphabetical listing cannot reveal. Theories of lexical semantics try to take account of these multiple interrelations between and among lexicalized concepts.

### *Lexical Semantics*

The English lexicon can be divided into closed-class words and open-class words. Closed-class words include all the little words—articles, prepositions, pronouns, conjunctions, and the like; membership in these classes changes slowly. Closed-class words generally play a grammatical role. Lexical semantics is concerned primarily with open-class words—the nouns, verbs, adjectives, and adverbs; these words can play a referential role and are always open to new members. So there is a creative aspect to the lexicon. Not only can new open-class words be coined to express new concepts, but familiar open-class words can also be used to express new concepts.

Two facts about open-class words are immediately apparent: There are a great many of them and their meanings are intricately interrelated. These two facts could hardly be unrelated. People would not be able to master so many

different words if they were all totally independent of one another. But one of the central problems of lexical semantics is to make explicit the relations between lexicalized concepts.

Essentially, two ways of analyzing word meanings have been proposed, relational and compositional. Relational lexical semantics was first introduced by Carnap (1956) in the form of meaning postulates, where each postulate stated a semantic relation between words. A meaning postulate might look something like  $\text{dog} \rightarrow \text{animal}$  (if  $x$  is a dog then  $x$  is an animal) or, adding logical constants,  $\text{bachelor} \rightarrow \text{man}$  and  $\text{never married}$  [if  $x$  is a bachelor then  $x$  is a man and  $\text{not}(x \text{ has married})$ ] or  $\text{tall} \rightarrow \text{not short}$  [if  $x$  is tall then  $\text{not}(x \text{ is short})$ ]. The meaning of a word was given, roughly, by the set of all meaning postulates in which it occurs.

Compositional lexical semantics, introduced by Katz & Fodor (1963), analyzes the meaning of a word in much the same way a sentence is analyzed into semantic components. The semantic components of a word are not themselves considered to be words, but are abstract elements (semantic atoms) postulated in order to describe word meanings (semantic molecules) and to explain the semantic relations between words. For example, the representation of “bachelor” might be ANIMATE and HUMAN and MALE and ADULT and NEVER MARRIED. The representation of “man” might be ANIMATE and HUMAN and MALE and ADULT; because all the semantic components of “man” are included in the semantic components of “bachelor,” it can be inferred that  $\text{bachelor} \rightarrow \text{man}$ . In addition, there are implicational rules between semantic components, e.g.  $\text{HUMAN} \rightarrow \text{ANIMATE}$ , which also look very much like meaning postulates.

These two ways of analyzing entailments between words appear different at first, but the semantic relations they deal with turn out to be similar. It may even be the case that any analysis formulated in one theory can be translated into the other. In either case, the interrelations between words take the form of inferences: if something is, say, a bachelor, one can infer that it is also animate and human and male and adult and unmarried. For that reason, mastery of these interrelations between words has been called inferential competence, in contrast to referential competence (Marconi 1997).

What is important is the semantic relations between words, not the notation in which those relations are represented. Consider a few of the more important semantic relations (Cruse 1986, Lyons 1977, Fellbaum 1998).

A basic semantic relation between open-class words is *synonymy*. The traditional way to define synonymy is in terms of substitution: Two words are synonyms (relative to a context) if there is a statement (or class of statements) in which they can be interchanged without affecting truth value. In natural languages there are few instances of synonyms that can be exchanged in every context. For example, “snake” and “serpent” can be interchanged in many con-

texts without altering truth values, but not in the statement, “The plumber used a snake to unclog the drain.” Another way to look at synonymy is as the extreme on a scale of semantic similarity.

Parallel to synonymy is the semantic relation of *antonymy*. Antonymy is sometimes used as a cover term for all kinds of lexical opposites, and sometimes used in a narrower sense to describe only the gradable oppositions where there is a neutral area between them. For example, “It is hot” implies that it is not cold, but “It is not cold” does not imply that it is hot, since there is a tepid range separating the antonyms “hot” and “cold.”

*Hyponymy* is a semantic relation describing the inclusion of one class in another. Semantic intuitions about hyponymy can be tested by sentences like *x is a y* or *x is a kind of y*; it is sometimes called the ISA relation (pronounced “is a”). For example, the acceptability of “A dog is a kind of animal” is interpreted to mean that “dog” is a hyponym of “animal.” Or in terms of class inclusion, considering that the class of animals includes the subclass of dogs, “dog” is a hyponym of “animal” [and “animal” is a hypernym (or superordinate) of “dog”]. Hyponymy generates hierarchies (sometimes called trees, taxonomies, or ontologies), thus *dog* → *canine* → *carnivore* → *placental mammal* → *mammal* → *vertebrate* → *chordate* → *animal* → *organism* → *entity* (where → is read “is a kind of”).

A similar semantic relation between verbs, *troponymy*, can be tested by sentences like *x-ing is a manner of y-ing*. For example, if “marching is a manner of walking” is acceptable, then the verb “march” is a troponym of the verb “walk.”

The semantic relation between words denoting parts and wholes is *meronymy*, which can be tested with such sentences as “The wrist is a part of the arm” or “An arm has a wrist,” indicating that “wrist” is a meronym of “arm” (and “arm” is a holonym of “wrist”). Sometimes meronyms can be arranged in a hierarchy, as in the case of body parts. But there are also meronyms like “handle,” which is a part of many things—hammers, umbrellas, briefcases, teacups, saucepans, and so on—so the hierarchy will usually be tangled.

Many other semantic relations can be described (Mel’cuk & Zholkovsky 1988), but the ones listed here are the relations that occur most broadly throughout the lexicon and are the most familiar. Even with this limited list, however, it is apparent that there is an interaction between syntactic categories and semantic relations. Every open-class syntactic category—nouns, verbs, adjectives, adverbs—has synonyms and antonyms, but hyponymy and meronymy are limited to nouns, troponymy to verbs.

When lexical semanticists assume that these relations can be verified by interchanging words in sentences or by the acceptability of sentences like *x is a kind of y*, or *x-ing is a manner of y-ing*, or *x is a part of y*, they are assuming that these semantic relations are part of what every native speaker knows about the



words of the language. And if Carnap were right—if the meaning of a word were simply the set of semantic relations it occurred in—then this would be the answer we seek. What a person knows when they know a word would then be simply given by its position in a vast net of words connected by these various syntactic and semantic relations (Quillian 1968, Miller 1978).

Psychologists, of course, want a cognitive theory of this vast semantic network, so exploring the network soon became a topic for psychological experimentation. (But for reasons never made clear, the initial investigations tended to concentrate on nominal concepts and to ignore words in other syntactic categories.) In 1969, Collins & Quillian published a seminal paper claiming that distances in the net could be inferred from the times people took to verify various test sentences. For example, “A canary is a bird” can be verified faster than can “A canary is an animal.” Smith (1978) summarized the outburst of experimentation stimulated by this proposal. In short, many exceptions were quickly discovered, and no single theory seemed to explain all the diverse verification times that were observed. One possible conclusion might have been that measuring verification times and eliciting judgments of semantic similarity are not appropriate methods for exploring semantic relations, but instead most psychologists rejected the idea that lexical memory contains a hierarchy of lexicalized nominal concepts.

Even if verification times had proved to be reliable indicators of semantic relations, the result would have been far short of a satisfactory cognitive theory of lexical knowledge. A person who knows a word knows far more than its semantic relations to other words. But a semantic account is not an unreasonable place to begin a search for a cognitive theory.

### *WordNet*

The basic structure of any lexicon is a mapping between word forms and word meanings. If the word forms are alphabetized, it is usually called a dictionary. A modern dictionary, however, provides far more information than would a simple enumeration of word meanings: spelling, pronunciation, etymology, part of speech, inflectional forms, derivational forms, sentences or phrases illustrating the use of the word to express different meanings, occasional pictures and occasional usage notes. Great patience and enormous scholarship are required for such work. Nevertheless, much of the information in a dictionary is familiar to native speakers of the language and so should probably be included in what a person who knows a word knows. But the semantic relations that are so important in lexical semantics are largely implicit in a conventional dictionary; they must be inferred from the definitional glosses.

It is hardly surprising that an alphabetical organization of lexical information should be totally unlike a cognitive organization. But the same information can, of course, be organized in many different ways. Beginning in 1985, a

group of cognitive scientists at Princeton University undertook to build a lexicon for a computer wherein the semantic relations between open-class words that are implicit in a dictionary could be made explicit in the computer (Fellbaum 1998). The result is an on-line lexical database called WordNet (Miller 1985). WordNet does not include pronunciations or etymologies or pictures or usage notes, but the basic form-meaning pairs are there.

WordNet is really four lexical databases, one each for nouns, verbs, adjectives, and adverbs. The basic building block of WordNet is a synset (a set of synonyms along with a standard dictionary-type definitional gloss); a synset is WordNet's representation of a lexicalized concept. If a word form appears in more than one synset, it is polysemous (WordNet treats homonymy as kind of polysemy).

Words and synsets are associated with one another by the familiar semantic relations. Synonymy and antonymy are lexical semantic relations between word forms in WordNet; hyponymy, troponymy, meronymy, and various verbal entailments are conceptual semantic relations between synsets. As this is written, WordNet contains 122,000 word forms in 100,000 synsets that are organized by 139,000 pointers representing semantic relations. Thus, the vast net of words that Quillian (1968) imagined has now actually been constructed in a computer.

WordNet has certain advantages over a machine-readable dictionary for the purposes of computational linguistics. For example, synonym sets can be used in information retrieval to expand a user's query and so retrieve relevant items that might otherwise have been missed. Again, many of the so-called semantic extensions of word meanings can be understood hyponymically; e.g. "board" can be used in different contexts to refer to a bulletin board or a drawing board or a surf board because hyponyms can be replaced by their superordinates. More interesting, it is possible with WordNet to retrieve lists of hyponyms; e.g. from the word "food" it is possible to retrieve the names of many different kinds of food. Thus, it is possible to go from, say, "eat food" to "eat x," where "x" is any of the different kinds of food, i.e. any hyponym of food. If x is a hyponym of food, "eat x" is semantically acceptable (Resnik 1993). Other applications of WordNet are described in Fellbaum (1998).

WordNet was constructed as an instantiation of hypotheses proposed in the 1980s. Its relevance to this discussion is to demonstrate what a lexical semantic system designed on those principles does not do. For example, one thing that WordNet does not do is to provide a topical organization of the lexicon. If someone were to ask for all the words used to discuss baseball, the bat and ball would be kinds of artifacts, the players would be kinds of people, the field would be a kind of location, and the game itself would be a kind of activity. Baseball words are scattered throughout WordNet, not clustered together by semantic relations. In short, a topical organization of a lexicon is very different

from a semantic organization. (And both, of course, are different from an alphabetical organization.)

Another thing WordNet does not do is provide a way to recognize the alternative meanings of a polysemous word. The alternative meanings of a polysemous word are distinguished, of course, as they are in conventional dictionaries, by synonyms and definitional glosses and, sometimes, by illustrative phrases or sentences. But nothing is said about using that information to decide whether an occurrence of the polysemous word in a particular context conforms to one meaning or another. WordNet is not unique in this neglect of contextual information. Dictionary definitions are deliberately decontextualized.

Lexicographers do not need to specify contexts of use in any detail because the people who use dictionaries seldom need such help. Native speakers are extremely good at recognizing the intended meaning of a polysemous word on the basis of the context in which it is used. But students learning English as a second language are familiar with this lexicographic deficiency; they look up an unfamiliar word in a dictionary where they may find half a dozen different definitions, but little help in determining which of those definitions is appropriate in the context of their reading.

It seems obvious, therefore, that native speakers must know something that is missing from dictionaries and thesauruses and lexical databases. Because the question at hand is what native speakers know when they know a word, this omission from theories of lexical semantics seems sufficiently serious to merit further discussion.

## CONTEXTUALIZATION

Although context of use is neglected in WordNet, it has played a major role in theories of many psycholinguistic phenomena. For example, the interactive social context in which children acquire their first language is so obviously important that it is normally taken for granted. Theories of early vocabulary development differ primarily in what kinds of constraints are proposed on children's appreciation of their interactive linguistic context.

Beyond the early years, reading begins to provide an alternative context for vocabulary growth. The rate at which school children learn new words has been estimated at around 20 words per day (Anglin 1993), which is far more than anyone could teach them. The usual explanation is that children learn words while reading, by observing the linguistic contexts in which they are used (Gray & Holmes 1938, Werner & Kaplan 1950, Sternberg & Powell 1983, Jenkins et al 1984, Nagy et al 1987). Unfortunately, learning words from context is a slow process. Many contexts of use must be encountered before a new word is mastered, so extensive reading is required for a large vocabulary.

Not only is context important for acquisition, it is critically important for comprehension. The dramatic example is deixis (Bühler 1934). If someone says, “I need you here now,” the referents of “I” and “you” are the speaker and a listener, who can only be identified from the actual context of use; that is deixis of person. Similarly, the referents of “here” and “now” depend on where and when the utterance occurred; those are deixis of place and time. These deictic words presuppose knowledge of the context in which they are used; they are meaningful, but they have no reference outside their context of use. Deixis illustrates how skillfully people use context.

Collocations also illustrate the importance of context. For example, “pen” has several meanings and so does “fountain,” but when they occur together in “fountain pen” all the possible combinations of meanings are ignored; the collocation has a single noncompositional meaning and can be treated as if it were a single word. The contribution of adjacent words is especially important for adjectives, whose meaning is strongly influenced by the nouns they modify: A long car and a long train are very different in length. “Long” and “short” express values of the attribute length, “wide” and “narrow” express values of the attribute width, and so forth, but the absolute values depend on the noun that is modified. “Good” and “bad” express values of the function that the head noun is normally expected to serve (Katz 1964): a good meal is nourishing, a good knife cuts well, a good friend is loyal, and other similar examples. The implication is that knowing the meaning of a noun entails knowing the attributes and functions of its referent.

Pustejovsky (1995) emphasized the generality of this phenomenon. One of his examples is the verb “enjoy,” where a missing predicate is inferred from information about the complement. In “She enjoyed the book” it is understood that she enjoyed reading the book because that is what one does with books; in “He enjoyed his coffee” it is understood that he enjoyed drinking his coffee because that is what one does with coffee, and so on. It must be part of the meaning of “book,” therefore, that one reads it, and part of the meaning of “coffee” that one drinks it. Thus, Pustejovsky would add information about attributes and composition and function and agency to the definitions of some words, plus rules to infer the meaning of other words whose definitions were underspecified. Such an approach has many attractions, but until it has been fully worked out it will be difficult to evaluate.

Another example of the importance of context for comprehension is its use to avoid the potential ambiguities of polysemy.

### *Polysemy*

In developing scientific terminology, a deliberate attempt is made to avoid using the same symbol to mean two different things, but multiplicity of meaning

abounds under the less restrictive conditions of everyday discourse. It is a perverse feature of natural languages that the more frequently a word is used, the more polysemous it tends to be (Zipf 1945, Jastrezemski 1981).

It is useful to maintain a distinction between polysemy and ambiguity. Although polysemous open-class words are commonly assumed to be ambiguous, they are only potentially ambiguous. They are ambiguous in isolation, but open-class words are seldom used in isolation. In everyday usage, polysemous words are rarely ambiguous. People navigate daily through a sea of potential ambiguities, resolving them so easily that they seldom notice they are there. Consequently, the ubiquity of multiple meanings is easily overlooked, yet it provides clear proof, if proof is needed, of a remarkable human ability to cope with polysemy.

Psycholinguists have used polysemous words extensively in a variety of experiments attempting to test the hypothesis that components of the language comprehension process are autonomous modules (Fodor 1983). If they were autonomous, then the occurrence of a polysemous word should lead to the simultaneous activation of all of its different meanings, and that information would then be passed as input to a separate module that would select the meaning most appropriate to the given context. If they were not autonomous, the activation of meanings and the evaluation of context could interact, and only the context-appropriate sense would be activated. The issue is complex, however, because any failure to obtain the predicted results could mean either that meaning identification is not modular or that the modules involved have been incorrectly identified. Simpson (1994) gives an excellent review of these studies, which provide strong evidence on both sides of the question; he suggests that the results must be highly dependent on characteristics of the context and characteristics of the tasks required of the subject. He concludes, "Only by taking a broader view of context will we be able to direct our efforts more productively to understanding the nature of lexical ambiguity and its relation to other aspects of language comprehension, rather than focusing on it exclusively as a means to settle a single theoretical debate" (Simpson 1994: 372).

The cognitive processes studied in these experiments occur in a fraction of a second, so fast that in normal listening or reading the alternative meanings, even when they are activated, do not rise above some threshold of conscious awareness. This speed is essential, of course. If polysemy could not be quickly resolved, comprehension would be swamped by a backlog of unprocessed semantic information. Cross-modal priming experiments have established that semantic processing can begin even before a spoken word is recognized (Marslen-Wilson 1987, Moss et al 1997), presumably with contextual facilitation. Thus, the experiments show that the intended sense is identified very quickly, but they say little or nothing about how this trick is performed.

### *Contextual Representations*

Simpson is surely right that a broader view of context is needed (Simpson 1994). A contextual approach to lexical semantics might assume that the meanings of a word form “are fully reflected in appropriate aspects of the relations it contracts with actual and potential contexts” (Cruse 1986:1). The problem is to determine what those “appropriate aspects” are. The lexical semantic relations discussed above must be part of the answer, but it is already clear that something more is required. Miller & Charles (1991) have given it a name: What is needed is a contextual representation—a cognitive representation of the set of contexts in which a given word form can be used to express a given word meaning.

To focus on the context rather than the polysemous word itself requires something of a figure-ground reversal. Charles (1988) introduced a sorting method in which context becomes the object of central interest. He selected two different words, then searched a textual corpus to find a haphazard sample of sentences using those words. Then he printed each context on a separate card, leaving a blank where the target words had been. He shuffled the deck of cards and handed it to subjects with the request to “sort them into piles so that the same word has been deleted from every sentence in the same pile.” In other words, Charles asked subjects to discriminate between the contexts of various words; he applied signal detection theory to his results in order to obtain values of  $d'$  as a measure of contextual discriminability.

His first discovery was that contextual discriminability seems to be inversely related to the semantic similarity of the deleted words. The larger the  $d'$ , the smaller the semantic similarity (Miller & Charles 1991). With words like “car” and “automobile” there was considerable confusion of contexts, but when semantically unrelated words like “noon” and “string” were used, their contexts were almost never confused. Miller & Charles proposed that semantic similarity might be assessed in terms of the relative number of contexts in which two words could be interchanged without creating a semantic anomaly.

Miller & Charles (1991) summarized a series of such experiments with the generalization that, associated with each word meaning, there must be a contextual representation. In particular, a polysemous word must have different contextual representations for each of its different senses (Miller & Leacock 1998). A contextual representation is not itself a linguistic context, but is an abstract cognitive structure that accumulates from encounters with a word in various linguistic contexts and that enables the recognition of similar contexts as they occur.

Note that contextual representations are precisely what is missing from most dictionary definitions. But it is not easy to explain to lexicographers what more they should provide. Unfortunately, “contextual representation” is not an explanation; it is merely a name for the thing that needs to be explained.

One thing is clear, however. A lexical representation must contain information limiting the number of alternative meanings that a polysemous word can have. Indeed, a general feature of contextual representations seems to be that they enable people to narrow the range of relevant alternatives they must cope with.

Three sources for the kinds of information embodied in contextual representations (Miller 1978) can be mentioned: situational context, topical context, and local context.

By *situational context* is meant the kind of information required for deixis or, more generally, information about the purposes and goals of the communicative interaction. It is generally assumed that an appreciation of the situational context draws heavily on a person's general knowledge of people and the world they live in. How world knowledge should be characterized is a controversial question that cannot be settled here. But one kind of situational context might be open to observation.

The situational context created by instructions given to participants in psychological experiments should be of particular interest. Giving instructions is an experimenter's way of providing a context that makes an experimental situation meaningful. In the 1950s when information theory (Shannon 1948) was first making psychologists aware of the importance of specifying the number of alternative stimuli and responses in their experiments, it was observed in a variety of experimental situations that instructions could determine what a subject expected, and those expectations could determine performance. That is to say, given instructions to do so, people are very skillful at restricting their attention to a limited range of possible alternatives.

*Topical context* is another important source of contextual information that can limit the interpretation of polysemous words. For example, the polysemous noun "shot" can have different referents depending on the domain of discourse in which it is used. In a discussion of marksmanship, shot means one thing; to a bartender it is something else; a doctor or nurse gives it with a hypodermic needle; to a photographer it is a snapshot; golfers and basketball players take still different shots. Each different meaning is associated with a different topic. So if the topic could be identified from the linguistic context, it might be possible to tell which meaning was intended.

Looking at topical context is an old idea (Kintsch 1972), although it is not clear exactly what a topic is or how people are able to identify one. The general notion is that each topic has its own vocabulary, so the problem is to learn what those vocabularies are, and what the meanings of polysemous words are in each different vocabulary. That knowledge is surely part of the contextual representations that help to determine intended meanings.

*Local context* refers to information provided by words in the immediate neighborhood of a polysemous word. Whereas topical context can disregard

word order, local context depends heavily on the order of the words and on their syntactic categories.

Knowing how to use a word implies knowing its syntactic category, so associated with every word form there must be a list of the syntactic roles it can play. Or, to reverse figure and ground again, there is one category of contexts in which nouns can appear, another category of contexts in which verbs can appear, and so forth. Some words can play more than one syntactic role, so to determine their syntactic category in any particular instance it is necessary to examine the context in which they are used. How people do this is an active research question, but one feature is apparent: The syntactic category of nearly every word can be determined immediately by looking at the other words in its local environment, without waiting for a full parse of the sentence.

The notion of a contextual representation, therefore, generalizes an idea already made familiar by syntactic categories. The generalization is that contextual representations are associated with word meanings, not word forms. Semantic as well as syntactic information is involved.

In 1955 Kaplan reported informal experiments in which he asked people to identify the meaning of a polysemous word when they saw only the words in its immediate vicinity. He found that they were almost always able to identify the intended meaning when they saw a string of five words, two to the left and two to the right of the target word. Kaplan's claim was welcomed by those working on machine translation, but those who found it easy to construct counterexamples were not impressed. Yet the result has been replicated (Choueka & Lusignan 1985) and appears to be correct. Local context is an important component of contextual representations.

It is one thing to recognize these sources of contextual information, but something else again to understand how they are used to identify the intended meanings of polysemous words in context. One approach is to simulate sense identification with a computer.

### *Automatic Word-Sense Identification*

Although word-meaning identification is no problem for people, it is a serious problem for computers. It is a major stumbling block for natural language processing by machine.

The difficulty was recognized during the early enthusiasm for machine translation in the 1950s and 1960s. Bar-Hillel (1960) concluded that fully automatic machine translation would be impossible because he believed that automatic word-sense identification is impossible. Word-sense identification is so important, however, that attempts to solve it continued without interruption (Hirst 1987).

An approach that has been popular recently takes advantage of the statistical learning systems and the large machine-readable corpora that are now



available. For example, Gale et al (1992) described a statistical approach that took advantage of topical context. They partitioned a sizable number of occurrences of a polysemous word into sets illustrating its different meanings, then, using a statistical classifier based on Bayesian decision theory, searched for co-occurring words that would be diagnostic of the topic and the appropriate sense of the target word. For example, the photographic sense of “shot” should occur with “lens” and “camera,” whereas “whiskey” or “brandy” should indicate another sense.

Leacock et al (1996) compared three different statistical learning systems using the noun “line.” They collected sentences containing line (or lines or Line or Lines) and sorted them according to meaning until they had a sizable sample of each of six different meanings. Then they divided the resulting sets of instances into training sets and test sets: The statistical sense-classifiers were trained on one set of contexts and tested on the other. For topical context they used a Bayesian classifier, a content vector classifier, and a neural net classifier. Results with the three statistical methods of topical classification were roughly equivalent (71%–76% correct meaning identification), although some meanings were consistently easier to recognize than others.

In order to check that the statistical classifiers were identifying contexts as well as possible on the basis of the information they were given, Princeton undergraduates were given lists of the co-occurring words (in reverse alphabetical order) and asked to guess which sense of “line” was intended. The students did only a little better than the statistical classifiers. The general similarity of their results was striking. Students had little difficulty with contexts that the classifiers got right, and had more trouble with the contexts that the classifiers got wrong. Leacock et al concluded that the statistical systems had reached a ceiling, that 20%–30% of the instances could not be classified on the basis of topical context alone.

Local context might be expected to be more helpful, but it is more difficult to extract and use. One way to exploit the information in local contexts is to use a template classifier (Weiss 1973). A template is a specific string of words (including the target word) that occurs frequently in the training contexts. For example, Leacock et al (1996) found that “a fine line between” correctly selected one meaning of line, whereas “wait in line” selected another, and “telephone lines” selected still a third. The results were excellent when the templates occurred, but all too often they failed to occur. Leacock et al (1998) had better success using a Bayesian method with local context.

There are many instances where the intended sense of a polysemous word can only be inferred by the use of situational context. How computers should be given information about the world and the goals of humans who live in that world is an important issue for artificial intelligence, but the results have not yet been brought to bear on the problem of automatic sense identification.

Automatic sense identification is an active area of research, and considerable ingenuity is apparent, but the results have been discouraging. The reality is that there are still no large-scale, operational systems for identifying the senses of words in connected text, and there probably will not be until nonlinguistic context can be taken into account. Indeed, it was my reluctant recognition of how difficult it is for computers to solve this problem that led me to wonder whether the human capacity to take advantage of context might not be a uniquely important cognitive process.

## CONCLUSION

Contextualization has been defined as the use of context to determine meaning and to resolve potential ambiguities. When contextualization is linked so closely to meaning, of course, it inherits all the uncertainties associated with the concept of meaning. But it also inherits the broad scope of meaningfulness. Wherever experience is meaningful, context must be considered. And language provides one of the best avenues to approach a study of the remarkable human capacity to use context.

Word-sense identification is a problem that must challenge anyone interested in how people use language to understand one another—how so much lexical polysemy can result in so little ambiguity. It is argued here that associated with every distinct meaning of a polysemous word there must be some cognitive representation of the contexts in which that word form can be used to express that meaning. But much remains to be done before the precise nature of these contextual representations will be understood.

Visit the *Annual Reviews* home page at  
<http://www.AnnualReviews.org>.

## Literature Cited

- Anglin JM. 1993. Vocabulary development: a morphological analysis. *Monogr. Soc. Res. Child Dev.* 58 (10), Ser. 238:1–166
- Atkins BTS. 1992/1993. Theoretical lexicography and its relation to dictionary-making. *Dictionaries: J. Dict. Soc. North Am.* 14:4–39
- Bar-Hillel Y. 1960. A demonstration of the nonfeasibility of fully automatic high quality translation. In *Advances in Computers*, ed. FL Alt, 1:158–63. New York: Academic
- Beck I, McKeown M. 1991. Conditions of vocabulary acquisition. In *Handbook of Reading Research*, ed. R Barr, LM Kamil, PB Mosenthal, PD Pearson, 2:789–814. New York: Longman
- Binet A, Simon T. 1911. *The Development of Intelligence in Children*. Transl. ES Kite, 1916. Baltimore, MD: Williams & Wilkins
- Bühler K. 1934. *Sprachtheorie: die Darstellungsfunktion der Sprache*. Jena: Gustav Fischer
- Carnap R. 1956. *Meaning and Necessity*. Chicago: Univ. Chicago Press. 2nd ed.
- Charles WG. 1988. The categorization of sen-

- tential contexts. *J. Psycholinguist. Res.* 17: 403–11
- Choueka Y, Lusignan S. 1985. Disambiguation by short contexts. *Comput. Humanit.* 19:147–57
- Collins AM, Quillian MR. 1969. Retrieval time from semantic memory. *J. Verbal Learn. Verbal Behav.* 8:240–47
- Cronbach LJ. 1942. An analysis of techniques for systematic vocabulary testing. *J. Educ. Res.* 36:206–17
- Cruse DA. 1986. *Lexical Semantics*. New York: Cambridge Univ. Press
- de Saussure F. 1916. *Cours de Linguistique Générale*. Transl. W. Baskin, 1959, in *Course in General Linguistics*. New York: Philosophical Library
- Fellbaum C, ed. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press
- Fodor JA. 1983. *Modularity of Mind*. Cambridge, MA: MIT Press
- Gale W, Church KW, Yarowsky D. 1992. A method for disambiguating word senses in a large corpus. *AT&T Bell Labs. Stat. Res. Rep.* 104
- Galton F. 1879. Psychometric experiments. *Brain* 2:149–62
- Gray W, Holmes E. 1938. *The Development of Meaning Vocabularies in Reading*. Chicago: Univ. Chicago Press
- Hirst G. 1987. *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge: Cambridge Univ. Press
- Jackendoff RS. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press
- Jastrezemski JE. 1981. Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cogn. Psychol.* 13:278–305
- Jenkins JR, Stein M, Wysocki K. 1984. Learning vocabulary through reading. *Am. Educ. Res. J.* 21:767–87
- Kaplan A. 1955. An experimental study of ambiguity and context. *Mech. Transl.* 2: 39–46
- Katz J. 1964. Semantic theory and the meaning of 'good'. *J. Philos.* 61:739–66
- Katz J, Fodor JA. 1963. The structure of a semantic theory. *Language* 39:170–210
- Kintsch W. 1972. *The Representation of Meaning in Memory*. Hillsdale, NJ: Erlbaum
- Landau SI. 1984. *Dictionaries: The Art and Craft of Lexicography*. New York: Scribner's
- Leacock C, Chodorow M, Miller GA. 1998. Using corpus statistics and WordNet relations for sense identification. *Comput. Ling.* 24: In press
- Leacock C, Towell G, Voorhees EM. 1996. Towards building contextual representations of word senses using statistical models. In *Corpus Processing for Lexical Acquisition*, ed. B Boguraev, J Pustejovsky, pp. 97–113. Cambridge, MA: MIT Press
- Lyons J. 1977. *Semantics*. Cambridge, UK: Cambridge Univ. Press
- Marconi D. 1997. *Lexical Competence*. Cambridge, MA: The MIT Press
- Marslen-Wilson WD. 1987. Functional parallelism in spoken word-recognition. *Cognition* 25:71–102
- Mel'cuk I, Zholkovsky A. 1988. The explanatory combinatorial dictionary. In *Relational Models of the Lexicon; Representing Knowledge in Semantic Networks*, ed. MW Evens, pp. 41–74. Cambridge, UK: Cambridge Univ. Press
- Miller GA. 1978. Semantic relations among words. In *Linguistic Theory and Psychological Reality*, ed. M Halle, J Bresnan, GA Miller, pp. 61–118. Cambridge, MA: MIT Press
- Miller GA. 1985. WordNet: A dictionary browser. *Information in Data, Proc. Conf. Univ. Waterloo Cent. New Oxf. Engl. Dict., 1st*, pp. 25–28, Waterloo, Ont., Can: Univ. Waterloo Cent. New Oxf. Engl. Dict.
- Miller GA. 1996. Contextuality. In *Mental Models in Cognitive Science: Essays in Honour of Phil Johnson-Laird*, ed. J Oakhill, A Garnham, pp.1–18. East Sussex, England: Psychology Press
- Miller GA, Charles WG. 1991. Contextual correlates of semantic similarity. *Lang. Cogn. Proc.* 6:1–28
- Miller GA, Johnson-Laird PN. 1976. *Language and Perception*. Cambridge, MA: Harvard Univ. Press
- Miller GA, Leacock C. 1998. Lexical representations for sentence processing. In *Polysemy: Theoretical and Computational Approaches*, ed. Y Ravin, C Leacock. New York: Oxford Univ. Press. In press
- Montague R. 1974. *Formal Philosophy: Selected Papers*. New Haven, CT: Yale Univ. Press
- Moss HE, McCormick SF, Tyler LK. 1997. The time course of activation of semantic information during spoken word recognition. *Lang. Cogn. Proc.* 12:695–731
- Nagy WE, Anderson RC, Herman PA. 1987. Learning word meanings from context during normal reading. *Am. Educ. Res. J.* 24:237–70
- Ogden CK. 1934. *The System of Basic English*. New York: Harcourt Brace
- Pustejovsky J. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press

- Quillian MR. 1968. Semantic memory. In *Semantic Information Processing*, ed. M Minsky, pp. 216–70. Cambridge, MA: MIT Press
- Resnik PS. 1993. *Selection and information: a class-based approach to lexical relationships*. PhD thesis. Univ. Penn., Phila.
- Seashore RH, Eckerson LD. 1940. The measurement of individual differences in general English vocabularies. *J. Educ. Psychol.* 31:14–37
- Shannon CE. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27: 379–423
- Simpson GB. 1994. Context and the processing of ambiguous words. In *Handbook of Psycholinguistics*, ed. MA Gernsbacher, pp. 359–74. San Diego, CA: Academic
- Smith EE. 1978. Theories of semantic memory. In *Handbook of Learning and Cognitive Processes: Linguistic Functions in Cognitive Theory*, ed. WK Estes. 6:1–56. Hillsdale, NJ: Erlbaum
- Sternberg RJ, Powell JS. 1983. Comprehending verbal comprehension. *Am. Psychol.* 38:878–93
- Thorndike EL. 1921. *The Teacher's Word Book*. New York: Teachers Coll., Columbia Univ.
- Weiss S. 1973. Learning to disambiguate. *Inform. Stor. Retr.* 9:33–41
- Werner H, Kaplan E. 1950. Development of word meaning through verbal context: an experimental study. *J. Psychol.* 29:251–57
- Zipf GK. 1945. The meaning-frequency relationship of words. *J. Gen. Psychol.* 33: 251–56