

## The mental representation of the meaning of words\*

P.N. JOHNSON-LAIRD

MRC Applied Psychology Unit,  
Cambridge

### Abstract

*Five phenomena concerning the meanings of words are outlined in this paper. They concern (1) our limited introspective access to the nature of lexical representations; (2) the existence of lexical entries that make accessible the sense of a word; (3) the effects of context on the interpretation of words; (4) the systematic gaps in the acquisition of lexical knowledge; and (5) the existence of different semantic types of open-class word. These phenomena are used as the basis for a psychological theory of meaning of words.*

### 1. Introduction

Outside a psychological laboratory, the recognition of words is seldom an end in itself, because listeners want to understand what they hear. Comprehension requires them to know and to retrieve the meaning of the words they recognize. Lexical meanings are the ingredients from which the sense of an utterance is made up, and its syntactic structure is the recipe by which they are combined. Listeners must put together the meanings of the words they recognize according to the grammatical relations that they perceive between them. Comprehension, however, does not end there, since it transcends purely linguistic knowledge. For example, anyone who knows English can retrieve the ingredients and combine them appropriately for a sentence such as:

Do you know who those people are?

The ingredients are the sense of the word “you”, the sense of the word “people”, and the senses of the other words in the sentence. But, the sense of the expressions must be distinguished from their reference—the particular entities or indi-

---

\*I am grateful to Lolly Tyler for forcing me to write this article and to Patrizia Tabossi and the referees for their cogent criticisms of it. Requests for reprints should be sent to P.N. Johnson-Laird, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom.

viduals that expressions pick out in the world. Reference from the standpoint of psychology is not merely a question of individuals in the real world: human beings invent imaginary and hypothetical worlds and regularly refer to individuals within them. Unlike certain logicians, ordinary people do not treat all expressions that refer to non-existent entities as equivalent.

To grasp the sense of a phrase such as “those people” is generally a precursor to determining its reference—the particular set of individuals to whom the speaker is referring in uttering the sentence. Grasping sense is a matter of knowing the language; determining reference is a matter of much more since it usually depends on knowledge of the situation, knowledge of the speaker, knowledge of the conventions governing discourse, and the ability to make inferences. In the absence of these components, no-one can go from the sense of a sentence to its real significance, which depends on who or what it is about and also on why the speaker uttered it. Listeners need to determine who is referred to by “you” and “those people” in the example above and whether the speaker is asking a simple question demanding only the answer “yes” or “no”, or making an indirect request for identifying information. They grasp the significance of the question only when they establish these facts.

There is, of course, no end to the process of recovering a speaker’s intentions. Listeners may infer that the speaker needs to identify the relevant people, they may infer why the speaker has that need, and so on. As the processing of speech proceeds from phonology through words to comprehension, it thus becomes increasingly dependent on inferences based on the social and physical circumstances of the utterance, on a knowledge of the situation to which it refers, and on general knowledge.

This article is about the mental representation of the meaning of words, but the inferential basis of the higher orders of comprehension must be borne in mind in trying to understand lexical semantics—if only because the major phenomena apply equally to the interpretation of both speech and writing.

The plan of the article is simple. It describes five phenomena that concern the mental representation of the meanings of words, that is, their senses, since their references depend on their contexts of use. These phenomena are important clues to how the mind represents meaning. After the description of these clues, they are used to motivate a theory of the mental representation of lexical meaning. Although the theory is driven by data—in much the same way that word recognition itself proceeds, the data were not collected as a result of theory-free observations. As many philosophers of science have emphasized, it is doubtful whether any observations can be made without at least the glimmerings of some theoretical motivation. In the present case,

however, the observations were made over a number of years and there is no simple unitary theory that led to them.

## **2. Consciousness and lexical meaning**

The single most obvious phenomenon about the meanings of words is the difficulty of focusing consciousness upon them. If I ask you what does the verb “sight” mean in the sentence:

He sighted a herd of elephants on the plain

then you are immediately aware that you know the meaning of the word, and that you understand the sentence with no difficulty. You should also be able to offer a paraphrase of the word, such as:

to see something at a distance.

But the formulation of this paraphrase is not an immediate and automatic process. You cannot turn to the appropriate definition in a mental dictionary and read out the contents that you find there. It may take a second or two to formulate a definition, and in some cases, as we shall see, you may be unable to give a helpful definition at all. In short, you have an immediate awareness of knowing the sense of a word, but you have no direct introspective access to the representation of its meaning.

The importance of this simple observation is twofold. First, it presents us with the problem that is the topic of this article, because if we had a ready access to lexical representations it would hardly be necessary to advance a theory about them. Second, the very distinction between what we can and cannot be conscious of constitutes an important clue to the architecture of the mind. A good theory of linguistic processing should explain why listeners can be aware of the words and intonation that speakers use, and aware of understanding (or not understanding) what the words mean. It should also explain why listeners lack immediate introspective access to the nature of the representations that underlie the meanings of words and sentences. An answer to this question will indeed be offered in the final section of the article.

## **3. The existence of lexical entries**

Because theorists are in the same predicament as everyone else when it comes to introspection, they lack any immediate evidence for the existence of a mental representation of the senses of words. Indeed, a major psychological

issue is whether there are lexical entries in the mind that give the meanings of words. Some theorists have assumed that the sense of a word consists of a structured set of semantic features into which its meaning is decomposed (e.g., Schaeffer & Wallace, 1970; Smith, Shoben, & Rips, 1974). Others assume that the mental lexicon takes the form of a semantic network (e.g., Anderson, 1976; Anderson & Bower, 1973; Collins & Quillian, 1969; Rumelhart, Lindsay, & Norman, 1972), or a combination of network and features (Glass & Holyoak, 1974/5). A third sort of theory, however, rejects the notion of semantic decomposition, and assumes that there are no semantic representations for words, only a vast set of rules of inference, or "meaning postulates" (see e.g., Fodor, Fodor, & Garrett, 1975; Fodor, 1977, Ch.5; Kintsch, 1974). Meaning postulates in such theories specify entailments that depend on words, for example

for any  $x$ ,  $y$ , if  $x$  is on the right of  $y$ , then  $y$  is on the left of  $x$ .

It is difficult to obtain crucial psychological evidence to decide amongst these theories. But, on the one hand, comprehension does not appear to call for a process of decomposition (see Fodor et al., 1975; Johnson-Laird, 1983); and, on the other hand, there is evidence which, though it was designed with another issue in mind, casts doubt on the meaning postulate theories (see Johnson-Laird, Gibbs, & de Mowbray, 1978). If readers wish to participate in a single trial of the experiment, which takes only a few minutes, they should carry out each of the following instructions without looking ahead to the next instruction.

(1) Scan as quickly as possible the list of words in Table 1; ticking in pencil those that denote things that are both solid and ordinarily fit for human consumption, for example, tick "pear", but not "whisky" which is consumable but not solid, and not "ivory" which is solid but not consumable. This is a simple task that ordinarily takes only a few seconds.

Table 1. *Search down these lists of words as quickly as possible for those that denote things that are normally solid (as opposed to liquid) and fit for human consumption*

sherry	knife	hammer	linoleum
ammonia	jug	ink	pippette
bucket	apple	cream	petrol
quartz	toaster	carafe	paraffin
skewer	syringe	needle	biscuit
broom	water	coal	plate
toffee	wood	veal	beer

- (2) Cover up Table 1 so that it is no longer visible.
- (3) Try to recall and to write down all the words in Table 1—every word, not just those that were ticked.

We carried out two experiments using a similar procedure, one in which the subjects listened to a brief auditory presentation of each word, and the other in which the subjects read through a list of words as quickly as possible. Both experiments showed that the more components that a word in the list had in common with the target category, the more likely it was to be remembered (see Table 2). Thus, a word such as "beer" which has one of the required components is more likely to be remembered than a word such as "petrol" which has neither of the key components. This result presumably reflects the amount of processing carried out on each word (Johnson-Laird et al., 1978; Ross, 1981), or the number of retrieval cues provided by the target components (McClelland, Rawles, & Sinclair, 1981), or both. It is neutral with respect to the existence of dictionary entries. However, Table 1 also includes words that denote, not substances, but utensils of various sorts. We found that such words in general were not so well recalled as the substance words, yet as Table 2 shows there was a significant trend within them. A word such as "plate" denotes a utensil that is used for consumable solids, whereas a word such as "vase" is used for non-consumable liquids. In general, the greater the match between the type of utensil and the target category the better the recall.

If there are no lexical entries but only a vast list of meaning postulates, subjects should reject all the utensils in the same way. Suppose the target category is "consumable solids", then they should search for postulates of the form:

For any  $x$ , if  $x$  is a  $\left\{ \begin{array}{c} \text{plate} \\ \text{jug} \\ \text{hammer} \\ \text{vase} \end{array} \right\}$  then  $x$  is consumable,

Table 2. *The percentages of words correctly recalled in the experiment carried out by Johnson-Laird, Gibbs, and de Mowbray (1978)*

	Semantic components of the target category possessed by the words			
	Both	One	Neither	Overall
Substance words	50.0	21.5	10.6	27.4
Utensil words	16.2	10.6	8.1	11.7

and fail to find them. Likewise, they would succeed in finding each of the postulates:

For any  $x$ , if  $x$  is a  $\left\{ \begin{array}{c} \text{plate} \\ \text{jug} \\ \text{hammer} \\ \text{vase} \end{array} \right\}$  then  $x$  is solid.

Hence the postulate theory cannot explain the trend in the data. However, if there are lexical entries from which the semantic information about a word is readily accessible, then the entry for a word such as “plate” will make available the fact that plates are utensils used to serve *consumable solids*, whereas the entry for “vase” will not make available any information containing these target components. Subjects searching the list for consumable solids are therefore likely to carry out more processing in order to reject “plate” than to reject “vase”, and this extra amount of processing accounts for the greater memorability of “plate”. A similar explanation in terms of the cues to recall provided by “consumable” and “solid” again depends on the ease of recovering the target components from the lexical entry for “plate”. The trend in the memorability for the utensil words can therefore be best explained on the assumption that there are comprehensive lexical entries containing specifications of the senses of words. The trend cannot readily be accounted for by inferences made after lexical access on the basis of independent meaning postulates.

#### 4. Context and lexical meanings

Linguistic context has well-known effects on the recognition of spoken and written words (see e.g. Fischler & Bloom, 1979; Meyer & Schvaneveldt, 1971; Schuberth & Eimas, 1977; Swinney, Onifer, Prather, & Hirshkowitz, 1979; Tanenhaus, 1987, this issue; Tweedy, Lapinsky & Schvaneveldt, 1977). It also has effects on the interpretation of words. This phenomenon is hardly surprising because words are notoriously ambiguous. There is considerable evidence which suggests that all the different senses of an ambiguous word are initially activated (Cairns & Kamerman, 1975; Conrad, 1974; Holmes, Arwas, & Garrett, 1977; Swinney, 1979). Yet the evidence may not be decisive. Patrizia Tabossi (personal communication) has made an interesting observation using the “cross-modal lexical decision task” developed by Swinney in which subjects hear a sentence and then at some point within it have to decide whether or not a visually presented string of letters is a word in the language. Tabossi found that where the disambiguating sentential context

brings to mind a salient aspect of the more frequent meaning of an ambiguous word within it, then the time to make the lexical decision is faster if the word relates to this salient feature than if it relates to the other meaning of the word. Thus, in Italian, the sentence:

Because of the terrible climate the explorers almost died in an expedition to the pole, which was long and difficult.

contains the ambiguous word, "polo", which may refer either to one of the world's poles or to the game played on horseback. The sentence not only disambiguates the word, but brings to mind a salient aspect of the world's poles, namely, their coldness. The time to decide that the string, "cold", presented visually immediately after the spoken word "polo" in the sentence, is reliably faster than the decision for the string, "horse", which relates to the other meaning of "polo". Some further results of Tabossi suggest that the effect does not arise from associative cueing by other words in the sentence. Perhaps contexts that bring to mind salient features of the main meaning of an ambiguous word eliminate the need to retrieve all of its meanings.

Listeners are normally aware of an ambiguity only if it is unresolved by the rest of the sentence. Hence the mechanism for resolving lexical ambiguities operates rapidly, automatically, and outside awareness. The standard linguistic account is that the mechanism centres on "selectional restrictions", that is, specifications of the senses of other words occurring in construction with the ambiguous word. Thus, the ambiguity of "board" is resolved in the sentence, "He sued the board", because the verb "sue" takes as its object only people and institutions: one cannot sue a plank of wood. This standard piece of lexicography was elevated into linguistic theory by Katz and Fodor (1963). Unfortunately, however, it has become clear that the crucial disambiguating component is often, not the sense of a word, but its reference. Consider the following discourse, for example:

The client received the cheque on Tuesday. He banked it.

The second sentence contains an ambiguous verb, which can be paraphrased as "to form a border or bank (out of some substance)", "to tilt (an aircraft) in flight", or "to deposit (money) in a bank". Yet, the sentence is unambiguous because the reference of "it" is plainly the cheque, and it is highly improbable that he formed a border or bank out of a cheque or was using it as an aircraft. For this and other reasons (Johnson-Laird, 1983, p. 233) it seems safer to assume that disambiguation generally depends on inferences based on a knowledge of the reference of expressions.

How many different meanings are there for a verb such as "eat"? Some linguists have argued that this verb, like many others, is highly polysemous.

Indeed, Weinreich (1966) claimed there are different senses of "eat" corresponding to eating soup with a spoon; eating a steak with a knife and fork; eating chop suey with chopsticks, and so on. Halff, Ortony, and Anderson (1976) have similarly claimed that words have many meanings and that a particular sense of a word is "instantiated" when it is used in context. Thus, in the sentence, "A fish attacked a swimmer," the sense of "fish" that is instantiated is likely to be equivalent to "shark". Anderson and his colleagues have reported a number of experiments in which a word corresponding to an instantiation, for example "shark", turns out to be a better recall cue to the sentence than the original word that occurred in it, e.g. "fish" (see Anderson & Ortony, 1975; Anderson, Pichert, Goetz, Schallert, Stevens, & Trollip, 1976). Garnham (1979) has obtained the same effect with verbs, for example "fried" is a better recall cue than the original verb for "The housewife cooked the chips", though not, as is to be expected, for "The housewife cooked the peas".

In fact, there has been too much emphasis on polysemy and in consequence a mistaken view about the mechanism of instantiation. Linguists have formulated more accurate linguistic criteria for ambiguity (Zwicky & Sadock, 1973), and the crucial psychological criterion is whether or not it is necessary to postulate more than one semantic representation for a word in order to account for the interpretations of the sentences in which it occurs. Instead of asking how many different meanings can be squeezed out of the word, psycholinguists need to ask what is the minimum number of different senses that are necessary to cope with all of its different uses. If "eat" were truly polysemous then the sentence:

He eats the food

should be highly ambiguous. It should have many wholly distinct senses. Yet it remains unequivocal. What is true, however, is that the sentence in common with others can be truthfully asserted of an infinite number of different situations: "he" can refer to any male individual, "food" can designate an indefinite number of different types of food served in an indefinite number of different conditions, and the manner by which the food is eaten can vary in an indefinite number of different ways from chewing it like cud to straining it through the teeth. This indeterminacy of reference is not sufficient to establish ambiguity because, if it were, all open-class words would be infinitely ambiguous and their meanings could not be contained by a finite brain. Hence the sentence above, which truly applies to a variety of situations, is referentially indeterminate, but not ambiguous. Its syntax is unambiguous, and its words are unambiguous: they each have in ordinary usage a single sense, but these senses suffice, as do the senses of all words, to embrace many different



situations. The sentence requires only a single representation of its meaning.

A comparable mistake has been made in the standard interpretation of instantiation. Context can, of course, pick out the appropriate sense of a genuinely ambiguous word, for example "He banked the cheque." However, the instantiation of an unambiguous word such as "fish" by a sentential context does not depend on picking out one sense from a long list of possibilities. A simple thought experiment, which was proposed in Johnson-Laird (1981), suggests a more plausible interpretation. Consider the sentence:

It frightened the swimmer.

It may well be that the word "shark" would make a better recall cue for this sentence than the original word, the pronoun "it", that functions as its subject. However, it is obvious that this pronoun does not have a vast set of different senses: it has a single sense that enables it to refer to any of a potentially infinite set of entities. Its reference can depend on its linguistic context if it is used to refer to something that is identified elsewhere in the discourse, or it can depend directly on the reference situation if it is used deictically. Instantiation is therefore a process, not of eliminating senses from a list in a lexical entry, but of imagining a more specific situation than is warranted solely by the meanings of words (see also Gumenik, 1979, for results that can be interpreted in the same way).

All open-class words, such as "fish" and "eat", are closer to being pronouns than is commonly recognized: they provide a relatively simple semantic framework that can be enriched by inferences based on knowledge. These inferences concern the situation designated by the sentence, and different linguistic contexts highlight different aspects of lexical meaning. Consider, for instance, the following sentences:

The tomato rolled across the floor.

The sun was a ripe tomato.

He accidentally sat on a tomato.

The first sentence calls to mind the characteristic shape of a tomato, the second its characteristic colour, and the third its characteristic squashiness (see Johnson-Laird, 1975). Listeners know all these aspects of tomatoes, and many more, but when they initially interpret a sentence they are most unlikely to call to mind all of this information (*pace* Gibson, 1971) or none of it (*pace* Fodor, Fodor, & Garrett, 1975). Instead, they are likely to retrieve some information—the most relevant for imagining the state of affairs depicted by the sentence, and the rest of the sentence is one obvious cue to what is relevant.

This hypothesis has been corroborated in a number of experiments carried

out by the author and his colleagues. Thus, the occurrence of a verb such as “pleased” suggests that the object of the sentence will be something that is animate, and subjects are indeed faster to detect the presence of an animate noun when it occurs in such a sentence than when it occurs in a sentence with a verb such as “soaked” (see Hodgkin, 1977). The facilitation occurs even when the target noun occurs prior to the verb. Similarly, if subjects are asked a specific question that hinges on the sense of a word, such as:

Is a diamond brilliant?

then, as Tabossi and Johnson-Laird (1980) have shown, their response is faster when the question follows a sentence such as:

The mirror dispersed the light from the diamond

than when it follows a sentence that does not call to mind the relevant aspect of diamonds:

The film showed the person with the diamond.

As we expected, subjects are slower to answer the question when the preceding sentence calls to mind some other but irrelevant aspect of diamonds, such as their hardness:

The goldsmith cut the glass with the diamond.

Table 3 presents the mean latencies to respond correctly to the questions and the mean numbers of errors. Subsequent experiments have shown that the phenomenon is equally apparent whether the priming is a result of selectional restrictions on the sense of a word or factual inferences about its likely referent (Tabossi, 1982).

For all of these experiments, independent panels of judges established that the priming sentences genuinely called to mind the relevant element of meaning, and the design made it very difficult for the subjects to guess which word in a sentence the subsequent question would be about or what the question

Table 3. *The mean latencies (ms) and mean errors (max = 12) to respond to questions about nouns in the Tabossi and Johnson-Laird (1980) experiment*

	Responses after a relevant priming sentence	Responses after a non-priming sentence	Responses after an irrelevant priming sentence
Latencies	1016	1089	1142
Errors	0.54	0.88	1.33

would be. Tabossi (1983) has even shown that there is a more general form of priming in which the initial sentence need not contain any of the nouns in the question. Thus, for example, the sentence:

The fountain pen left a spot in the desk drawer

enables subjects to respond faster to the subsequent question:

Does ink leave a stain?

than when it occurs after a neutral sentence that does not call to mind the relevant property of ink.

Linguistic context evidently has at least three different effects on the interpretation of words. First, it can enable the listener to select the appropriate sense of a truly ambiguous word. Second, it can lead to a representation of more specific referents than is strictly warranted by the sense of an unambiguous word. For example, a listener imagines a shark as an instance of the generic term, “fish”, since a shark is a plausible actor in the situation described by the sentence. Third, it can call to mind particular aspects of a word’s interpretation—at the expense of other aspects. Thus, it plays a major part in the interpretation of compound nouns, such as “hot dog man” (see Clark, 1983). The context of a cooperative game can even lead people to a tacit negotiation of specific meanings for general nouns, such as “row” and “column” (Anderson, 1983). What has sometimes been underestimated in all of these cases is the importance of reference, or more properly, its psychological correlate: the representation of specific referents, real or imaginary, in particular situations. What the context refers to can disambiguate a word; it can instantiate a more specific referent; and it can suggest an aspect of a word’s meaning that is particularly relevant to what is going on.

## **5. The acquisition of lexical meanings**

People often do not know the meaning of a word in their language. Such ignorance may not matter. If someone says:

The explorers survived on pemmican and truffles

you may readily understand this remark, and only on being specifically questioned realize that you do not know exactly what pemmican and truffles are. The reason that an incomplete grasp of lexical meaning may be sufficient for comprehension is that you are nevertheless able to imagine the state of affairs described by the sentence. The evidence of the previous section shows that you do not necessarily retrieve all the semantic information that you possess

about a word. If you lack some information, the gap may go unnoticed where it is not crucial to understanding the sentence.

Gaps in lexical knowledge are predictable. People are likely to be aware of what is important, and thus, for instance, if they know anything about the sense of a word they should know whether or not it means a substance fit for human consumption. They are similarly more likely to be aware of a perceptible property, such as whether a substance is solid or liquid, than of a more covert property, such as its provenance (whether it is natural or manmade). Graham Gibbs and I quizzed two groups of subjects about these three aspects of a set of rare words (see Johnson-Laird, 1975). Typically, our subjects knew for instance that “pemmican” was consumable and that “verdigris” was not, but their knowledge of the structure and provenance of these substances was less secure. Table 4 presents the mean numbers of errors that the subjects made on a set of 48 rare words. The trend was reliable for both groups. Of course, exceptions to the general trend are to be expected where a particular aspect of a substance is highly salient, and such exceptions have been demonstrated by Emma Coope (in an unpublished experiment).

Gaps in lexical knowledge point to the importance of the process of acquisition, since the way in which concepts are acquired will inevitably be reflected in the form and content of lexical entries. There are two obvious processes by which you can acquire the meaning of a word: you can be told what the word means or you can infer what it means from encountering it in use. To be told the meaning of a word presupposes that it is possible to frame a useful definition of its meaning. Jerry Fodor has often claimed that there are no good definitions (see e.g., Fodor, Garrett, Walker, & Parkes, 1980). The truth is—as many lexicographers would assert—there are no good definitions for *some* words. For other words, there are excellent definitions. Indeed the majority of words in the Oxford English Dictionary can only be acquired by definition because they hardly ever occur in actual discourse. Such words are in fact easy to define in a way that is genuinely informative, e.g. “an arblast

Table 4. *The mean errors in categorizing 48 rare words on three semantic contrasts*

	Sample 1: University students ( <i>N</i> = 24)	Sample 2: Technical college students ( <i>N</i> = 12)
Consumable/Nonconsumable	4.7	5.0
Solid/Liquid	6.7	7.3
Natural/Manmade	9.1	10.0

is a cross-bow, consisting of a steel bow fitted to a wooden shaft, furnished with a special mechanism for drawing and letting slip the bowstring, and discharging arrows, bolts, stones, etc." Other words, however, are singularly difficult to define in a way that is useful. Dr. Johnson was perhaps satirizing the futility of definition in these cases when he defined a network as "anything reticulated or decussated, at equal distances, with interstices between the intersections". Anyone who does not know the meaning of the definiens is hardly likely to be helped by the definiendum.

Is there any way of predicting the difficulty of defining the meaning of a word? Gerry Quinn and I set out to answer this question in an experimental study of definitions. We asked our subjects to try to define a series of verbs in a way that would help children or foreigners whose grasp of English was insecure. We chose four levels of semantic complexity of the verbs following the analyses of Miller and Johnson-Laird (1976), and we predicted that semantically complex verbs, such as "watch" and "lend", would be easier to define than the semantically simplest verbs, such as "see" and "own". It should be easy to break down the meaning of a complex verb into simpler components for which there are corresponding words, but it should be hard to find such components for a simple verb. Our prediction was confirmed. For the simplest of the verbs, the subjects could at best offer only synonyms, which would not be very helpful to poor speakers of the language. As for the remaining verbs, the more complex they were, the easier the subjects found the task and the more accurate their definitions (see Johnson-Laird & Quinn, 1976).

The traditional account of lexical acquisition is that a child learns an association between a word and the thing that it denotes. There are many problems with this idea—establishing the set of referents for a word should not be confused with the mere conditioning of a stimulus (Harrison, 1972), the word could designate any of the manifold properties of the object rather than the object itself (Wittgenstein, 1953), and many words have either no perceptible referent or else are parts of speech for which the notion is irrelevant. Above all, however, children are no mere passive receivers of word-object associations: they entertain their own hypotheses about the meanings of words (Bowerman, 1977), and they coin their own words if no-one provides them with a suitable term (Clark, 1982). Hence, although children acquire words from observing them in use, a comprehensive theory of this process, such as might be modelled in a computer program, is a long way from being formulated. There are even theorists who are so perplexed by the process that they argue that learning has little role to play in it, and that concepts are innate and merely "triggered" by experience (Fodor, 1980). Although a native endowment is crucial, the phenomena above and some that I will describe in a

moment imply that a form of learning does underlie the acquisition of lexical meanings.

Conservative estimates of the rate at which children acquire words suggest that at around the age of five they are adding to their vocabulary some 20 or more words per day (see e.g., Templin, 1957; Miller, 1977, 1986). So rapid a rate is hardly consistent with a theory that allows only for simple associative learning. One interesting conjecture is that children can pick up elements of the meaning of a word merely from hearing it used appropriately in constructions containing words that they already understand. Til Wykes and I confirmed this conjecture in an experiment with 3- and 4-year-olds. The children listened twice to a series of stories. Each story contained a novel word that the children had not heard before. For example, one story featured the novel verb, "mib", which was used transitively with a meaning resembling "soak" and intransitively with a meaning resembling "spill"—our idea was to inhibit the children from merely substituting a familiar synonym for the nonsense syllable. After they had heard the story twice, the children were able to pick out the one entity (orange juice) that could mib from a set of four alternatives. Their performance was similar for the other three nonsense verbs, and it remained above chance one week later when they had to carry out the same task with a new set of alternatives (see Wykes & Johnson-Laird, 1977). In an unpublished study, Jon Davies showed that children could also acquire elements of the meanings of nonsense *nouns* from hearing them used in constructions with verbs with which they were familiar.

There may be an analogy between acquiring a language and the implementation of a compiler for a new high level programming language. A compiler is a special program for taking programs written in the new language and translating them into the machine code that controls the actual operation of the computer. It is sensible to write part of the compiler in assembly language (which maps readily into machine code), and to write the rest of the compiler in the new language itself. The former translates the latter into machine code, and saves the designers from the chore of writing the whole of the compiler in assembly language. It is not too far-fetched to imagine that lexical learning lifts itself up by its own bootstraps in a similar way. Children first learn, or perhaps know innately, how to relate certain internal representations to states of affairs in the world. Once a fragment of the language has been mapped onto this knowledge, it becomes possible to acquire other words indirectly by inferring their meaning from the contexts in which they occur or by being given explicit definitions of them. Some words are likely to fall clearly into the category of those acquired by direct acquaintance, for example simple words like "see" and "own" that are so hard to define; other words are likely to fall clearly into the category of indirect acquisitions, for example "arblast"

and “eleemosynary”. Many words, however, will be of mixed acquisition; and different individuals will acquire a given word in different ways.

## **6. Meanings and prototypes**

To understand an assertion is to know how the world would be if the assertion were true. This formula does not imply that when you understand an assertion you know how to verify it, or indeed that it is possible to verify it. It is one thing to know how the world should be and quite another to know how to find out whether the world is in that condition. However, if you have no idea what constraints an assertion implies about reality, then you have no idea what it means. One striking feature of natural language is that for the language community as a whole there are lexical items (within the same syntactic category) that vary in the completeness with which their semantics specifies this information. Consider the earlier example:

He sighted a herd of elephants on the plain.

The function words and the words “sighted”, “herd”, and “plain”, have a complete semantics, because no conceivable advance in human knowledge can force us to add to our conception of their meaning or to cause us necessarily to modify it. The way we conceive the world given the truth of this utterance is, in principle, completely specified as far as the meanings of these words are concerned. The case is different for the word, “elephant”. Most speakers of English have a good idea of what an elephant is—they have seen an elephant, or a picture of one, and they know something of the nature of the beast. Yet the term is a theoretical one. It designates a set of creatures within our categorization of animals. Our knowledge of such matters is far from complete, and we are committed to the existence of the category without knowing for certain what the essentials of elephanthood actually are—indeed without knowing incontrovertibly that the class is a truly unitary one. Such words notoriously give rise to the problem of demarcating what should go into the dictionary from what should go into the encyclopedia—a problem for which there appears to be no principled solution (see Gerrig, 1985). These words are “natural kind terms”, and it is doubtful whether there are any necessary and sufficient conditions for defining them (Putnam, 1975).

The existence of natural kind terms has important implications for the contents of lexical entries. The entry for “elephant” is likely to include information that can be used for identifying elephants and for imagining them, as well as other conceptual information (Miller & Johnson-Laird, 1976). If I assert that I have sighted an elephant, then you will interpret my utterance

to mean that I saw a large, four-legged mammal with tusks and a trunk. Such interpretations cannot be mediated by meaning postulates or any other form of lexical representation that implies that these attributes are necessary components of elephants. They are not; an elephant may lack any of them. They are not essential characteristics; and they are not mere inductions, since to check them inductively presupposes some independent method of first identifying elephants. In fact, they are part of our "theory" of elephants, which tells us that a prototypical member of the class has each of these attributes.

Eleanor Rosch and her colleagues have collected much evidence that is consistent with the existence of prototypes (e.g., Rosch, 1976). Real objects, unlike many of the concepts studied in the psychological laboratory, have features that are correlated—if an animal has a trunk, it tends to have tusks—and such correlations will be reflected in the prototype. Likewise, not all instances of a concept are equally representative, and the speed with which instances are categorized depends on how prototypical they are (Rosch, 1973). The major problem with prototypes is how they are represented in the mental lexicon. Rosch (1976) has suggested that a prototype is represented by a concrete image of an average category member. Ironically, Kant (1787) had already raised a decisive objection to this theory:

In truth, it is not images of objects, but schemata, which lie at the foundation of our pure sensuous conceptions. No image could ever be adequate to our conception of triangles in general. For the generalness of the conception it could never attain to, as this includes under itself all triangles, whether right-angled, acute-angled, etc., whilst the image would always be limited to a single part of this sphere.

The lexical entry for "elephant" must therefore consist of a schema representing the prototypical animal, and perhaps the best way to think of a schema is in terms of a mental model defined in terms of an interrelated set of "default values" (Minsky, 1975), that is, specific values for variables that can be assumed in the absence of information to the contrary. Thus, default values have a different status to the normal representation of a word's contribution to the truth conditions of a sentence. Normal truth conditions support valid inferences since they are necessary components of a word's meaning. Default values place a weaker constraint on how the world should be: they hold only in the case that nothing is asserted to the contrary. Hence, your knowledge of the default values for "elephant" lead you to assume that I saw an animal with one trunk, two tusks, four legs, etc., unless you have evidence to the contrary.

Lexical entries containing default values still place constraints on the world, but they do so indirectly by way of the set of alternative prototypes



governing a domain. You will not necessarily judge that I spoke falsely if the animal I saw had no trunk and one tusk. But, you will think me mistaken if, on inspection, the beast turns out to fit the prototype of a rhinoceros, or alternatively not to fit the prototype of an animal at all.

## **7. Towards a theory of the representation of lexical meanings**

The clues of the five previous sections fit together to suggest a coherent picture of the meanings of words. This theory, which I will outline here, is intended to answer three central questions: What are the contents of lexical meanings? How are they mentally represented? And what is their role in speech and comprehension?

The evidence from the semantic search task implies that there are entries in the mental lexicon that allow ready access to the information that an individual has about the sense of a word. The contents of an entry may be incomplete in one of two distinct ways. First, the individual may have yet to acquire a complete semantics for the word; second, the word may be a theoretically based one for which there is only an incomplete semantics. There are other expressions and nonce words with meanings that depend essentially on the context in which they occur, for example, the verb “porched” as in “The newsboy porched the newspaper” (see Clark & Clark, 1979; Clark, 1983). Words that are acquired by direct acquaintance with their denotata are likely to have lexical entries that contain ineffable specifications of their truth conditions, that is, entries that specify how the world has to be for them to apply to it, and that are all but impossible to define. In the case of natural kind terms, a major component of the representation of sense will consist of default values.

Words with a more complex semantics may be acquired from verbal definitions, or from encountering their use in verbal expressions. Their lexical representation may accordingly relate them to other words. Most words in common usage are likely to possess elements of both sorts of information, for example, people have access to procedures for imagining elephants, and they have access to other conceptual information about elephants, which they may have acquired either from usage or from a definition, such as the fact that elephants are animals.

The theory therefore draws a basic distinction between ineffable truth conditions (akin to expressions in machine code) and verbal definitions (akin to expressions in a high level programming language). The distinction relates, of course, to the old arguments about the existence of semantic primitives. What it implies, however, is that although primitives exist they are remote

from the directly expressible analyses of the meanings of words. They are unanalysable by normal cognitive mechanisms, outside conscious awareness, and presumably innate. One can advance plausible conjectures, however, about the functions that they are used to compute, for example, the perceptual representation of the world, the representation of discourse in the form of an imagined model of the state of affairs it describes, and the choice of appropriate words to describe a perceived or imaginary state of affairs. Likewise, one can begin to advance hypotheses about their role in the identification of objects (Marr, 1982) and in the construction of mental models of discourse (Johnson-Laird, 1983).

The specifications of verbal relations in the lexicon can be based on some mechanism akin to a semantic network or to meaning postulates, though the power of such theories is likely to make it difficult to test them empirically (see Johnson-Laird, Herrmann, & Chaffin, 1984).

The specifications of truth conditions in the lexicon can be thought of as the ingredients necessary for the procedures that construct, modify, and manipulate mental models. Thus, the representation of, say, "on the left of" calls for a specification that will enable a verification routine to scan a mental model in the appropriate direction to verify the relation, and that will enable a construction routine to scan a mental model in the appropriate direction before adding an element to the model, and so on.

The specification of default values can depend on similar procedures, but their results in models can be undone in the light of other information. Exactly such procedures are needed in any case whenever a model is based on information that is insufficiently determinate to specify a unique situation, that is whenever a model is based on virtually any piece of discourse. I describe my office, for instance, and you form a mental model of the arrangement of the furniture, but since my description is bound to be consistent with more than one possibility, you may have to revise your model in the light of subsequent information (see Johnson-Laird, 1983, for a description of computer programs using both truth conditions and default conditions of these sorts).

The dichotomy between ineffable truth conditions and verbal formulae has a number of repercussions. The logical properties of words, for instance, can arise in two distinct ways: from a representation of an explicit verbal relation ("elephants are animals") or from the consequences of the representations of their truth conditions. Hence, if you know what has to be the case for something to be an elephant, and you know what has to be the case for something to be an animal, then a simple thought experiment will lead you to the same conclusion that elephants are animals. There are a number of clear cases where the logical properties of words arise only from their truth

conditions, because the vagaries of their logical behaviour are too great to be encompassed by simple verbal definitions, for example, natural language quantifiers, and spatial expressions such as "on the left of".

The contrast between verbal formulae and truth conditions also arises in the interpretation of discourse, which seems to call for a listener to construct an initial verbal representation close to the linguistic form of the utterance and then to use this representation, together with lexical entries, to construct a mental model of the discourse. Although the existence of these two levels of representation is a matter of controversy, they are borne out by the need for independent representations of sense and reference, by linguistic phenomena such as the two classes of anaphora (surface and deep), and by experimental results on the memory for discourse (see e.g. Mani & Johnson-Laird, 1982; Johnson-Laird, 1983).

A major problem confronting the present theory is to reconcile two important constraints on the process of comprehension. On the one hand, information from an utterance is integrated into the existing model as a function of the referential links, if any, between the utterance and the model; on the other hand, the interpretation of the sense of a sentence almost certainly depends on combining the senses of its constituent words according to the syntactic relations between them. No existing theory has yet shown how these two different demands can be met within a single unitary theory of comprehension.

One question remains: why do we lack a conscious access to the nature of lexical representations? The answer is that the truth conditions of words are intimately dependent on the mind's ability to relate representations to the world. There is a twofold evolutionary advantage in not having conscious access to such perceptual mechanisms: first, they can operate in parallel and therefore more efficiently; and, second, if you see a tiger, you take avoiding action rather than inspect the process of perception to ensure that it is operating correctly. The lexical system inherits the inaccessibility of this basic representational machinery. There is a further advantage in this lack of access: you do not become aware of a gap in lexical knowledge unless it is immediately germane to the interpretation of the discourse. If you had a conscious access to your lexical representations, then every time you encountered a word for which you possessed an incomplete semantics, you would be aware of it. You would be in a comparable state of mind to someone who looks up a word in a dictionary only to find that part of the relevant entry has been torn out. This intrusive awareness would occur even if the missing information were not actually required to construct a model of the discourse. Similarly, every time you encountered an ambiguous word, you would be aware of it—even if the ambiguity were resolved by the immediate context.

Since your aim is to grasp the significance of an utterance and perhaps to act upon it, your interpretative system has no need to present these details to consciousness, just as there is no need to make the details of the perceptual process accessible. The same consideration, of course, applies to the acquisition of meaning: children can acquire a new element of meaning *en passant* without becoming aware that they are so doing, and in this way they can attend primarily to the significance of the utterance rather than the process by which they are interpreting it.

## 8. Conclusions

The present theory of lexical meanings rests on seven principal assumptions:

(1) Comprehension requires the listener to construct a model of the state of affairs described by the discourse. Words contribute to the sense of utterances, but this model depends on inferences from context about the specific referents of expressions.

(2) There is a mental dictionary that contains entries in which the senses of words are represented.

(3) A lexical entry may be incomplete as a result of ignorance or because the word is a theoretical term with an intrinsically incomplete sense.

(4) The senses of words can be acquired from definitions or from encountering instances of the word in use. The former procedure can only work with words that contain a complex semantics.

(5) Corresponding to the method of acquisition, elements of a lexical representation can consist of (a) relations to other words, which could be represented by a mechanism akin to a semantic network, and (b) ineffable primitives that are used in constructing and manipulating mental models of the world.

(6) The primitive elements in a lexical representation may specify the word's contribution to the truth conditions of the expressions in which it occurs, or else the logically weaker default values of the word.

(7) The contrast between explicit verbal relations and ineffable truth conditions is related to the way in which discourse, in turn, is represented initially in a superficial linguistic form and subsequently in the form of a model of the state of affairs that it describes.

## References

- Anderson, A. (1983). *Semantic and social-pragmatic aspects of meaning in task-oriented dialogue*. Ph.D. Thesis, University of Glasgow.
- Anderson, J.R. (1976). *Language, memory, and thought*. Hillsdale, NJ.: Erlbaum.
- Anderson, J.R., & Bower, G.H. (1973). *Human associative memory*. Washington, D.C.: Winston.
- Anderson, R.C., & Ortony, A. (1975). On putting apples into bottles—a problem of polysemy. *Cognitive Psychology*, 7, 167–180.
- Anderson, R.C., Pichert, J.W., Goetz, E.T., Schallert, D.L., Stevens, K.V., & Trollip, S.R. (1976). Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior*, 15, 667–679.
- Bowerman, M. (1977). The acquisition of word meaning: an investigation of some current concepts. In N. Waterson and C. Snow (Eds.) *Development of communication: Social and pragmatic factors in language acquisition*. New York: Wiley.
- Cairns, H.S., & Kamerman, J. (1975). Lexical information processing during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 14, 170–179.
- Clark, E.V. (1982). The young word-maker: a case study of innovation in the child's lexicon. In E. Wanner and L.R. Gleitman (Eds.) *Language acquisition: The state of the art*. Cambridge: Cambridge University Press.
- Clark, E.V., & Clark, H.H. (1979). When nouns surface as verbs. *Language*, 55, 767–811.
- Clark, H.H. (1983). Making sense of nonce sense. In G.B. Flores d'Arcais and R. Jarvella (Eds.) *The process of understanding language*. New York: Wiley.
- Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240–247.
- Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory and Cognition*, 2, 130–138.
- Fischler, I., & Bloom, P. (1979). Automatic and attentional process in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 18, 1–20.
- Fodor, J.A. (1980). Fixation of belief and concept acquisition. In M. Piattelli-Palmarini (Ed.) *Language and learning: The debate between Jean Piaget and Noam Chomsky*. Cambridge, Mass.: Harvard University Press.
- Fodor, J.A., Garrett, M.F., Walker, E.C.T., & Parkes, C.H. (1980). Against definitions. *Cognition*, 8, 263–367.
- Fodor, J.D. (1977). *Semantics: Theories of meaning in generative grammar*. New York: Crowell.
- Fodor, J.D., Fodor, J.A., & Garrett, M.F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry*, 4, 515–531.
- Garnham, A. (1979). Instantiation of verbs. *Quarterly Journal of Experimental Psychology*, 31, 207–214.
- Gerrig, R.J. (1985). Process and products of lexical access. Unpublished MS., Department of Psychology, Yale University.
- Gibson, E.J. (1971). Perceptual learning and the theory of word perception. *Cognitive Psychology*, 2, 351–368.
- Glass, A.L., & Holyoak, K.J. (1974/5). Alternative conceptions of semantic memory. *Cognition*, 3, 313–339.
- Gumenik, W.E. (1979). The advantage of specific terms over general terms as cues for sentence recall: instantiation or retrieval? *Memory and Cognition*, 7, 240–244.
- Half, H.M., Ortony, A., & Anderson, R.C. (1976). A context-sensitive representation of word meanings. *Memory and Cognition*, 4, 378–383.
- Harrison, B. (1972). *Meaning and structure*. New York: Harper & Row.
- Hodgkin, D. (1977). *An experimental study of sentence comprehension and sentence meaning*. Ph.D. thesis, University of London.
- Holmes, V.M., Arwas, R., & Garrett, M.F. (1977). Prior context and the perception of lexically ambiguous sentences. *Memory and Cognition*, 5, 103–110.

- Johnson-Laird, P.N. (1975). In A. Kennedy & A. Wilkes (Eds.) *Studies in long term memory*. London: Wiley.
- Johnson-Laird, P.N. (1981). Mental models of meaning. In A.K. Joshi, B.L. Webber, & I.A. Sag (Eds.) *Elements of discourse understanding*. Cambridge: Cambridge University Press.
- Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press/Cambridge, Mass.: Harvard University Press.
- Johnson-Laird, P.N., Gibbs, G., & de Mowbray, J. (1978). Meaning, amount of processing, and memory for words. *Memory and Cognition*, 6, 372-375.
- Johnson-Laird, P.N., Herrmann, D., & Chaffin, R. (1984). Only connections: A critique of semantic networks. *Psychological Bulletin*, 96, 292-315.
- Johnson-Laird, P.N., & Quinn, J.G. (1976). To define true meaning. *Nature*, 264, 635-636.
- Kant, I. (1787). *The critique of pure reason*, Second edition, Translated by J.M.D. Meiklejohn, London: Dent (1934).
- Katz, J.J., & Fodor, J.A. (1963). The structure of a semantic theory. *Language*, 39, 170-210.
- Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum.
- Mani, K., & Johnson-Laird, P.N. (1982). The mental representation of spatial descriptions. *Memory and Cognition*, 10, 181-187.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- McClelland, A.G.R., Rawles, R.E., & Sinclair, F.E. (1981). The effects of search criteria and retrieval cue availability on memory for words. *Memory and Cognition*, 9, 164-168.
- Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227-234.
- Miller, G.A. (1977). *Spontaneous apprentices: Children and language*. New York: Seaburg Press.
- Miller, G.A. (1986). Dictionaries in the mind. *Language and Cognitive Processes*, 1, 171-185.
- Miller, G.A., & Johnson-Laird, P.N. (1976). *Language and perception*. Cambridge, MA: Harvard University Press/Cambridge: Cambridge University Press.
- Minsky, M. (1975). Frame-system theory. In R.C. Schank & B.L. Nash-Webber (Eds.) *Theoretical issues in natural language processing*. Preprints of a conference at MIT, Reprinted in P.N. Johnson-Laird and P.C. Wason (Eds.) *Thinking: Readings in cognitive science*. Cambridge: Cambridge University Press (1977).
- Putnam, H. (1975). The meaning of 'meaning'. In K. Gunderson (Ed.) *Language, mind and knowledge*. Minnesota Studies in the Philosophy of Science, Vol. 7. Minneapolis: University of Minnesota 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*. New York: Academic Press.
- Rosch, E. (1976). Classification of real-world objects: origins and representations in cognition. In S. Ehrlich & E. Tulving (Eds.) *La mémoire sémantique*. *Bulletin de Psychologie*, Paris. Reprinted in P.N. Johnson-Laird & P.C. Wason (Eds.) *Thinking: Readings in cognitive science*. Cambridge: Cambridge University Press (1977).
- Ross, B. (1981). The more, the better? Number of decisions as a determinant of memorability. *Memory and Cognition*, 9, 23-33.
- Rumelhart, D.E., Lindsay, P.H., & Norman, D.A. (1972). A process model for long term memory. In E. Tulving & W. Donaldson (Eds.) *Organization and memory*. New York: Academic Press.
- Schaeffer, B., & Wallace, R. (1970). The comparison of word meanings. *Journal of Experimental Psychology*, 86, 144-152.
- Schubert, R.E., & Eimas, P.D. (1977). Effects of context in the classification of words and non-words. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 27-36.
- Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and process in semantic memory: a featural model for semantic decisions. *Psychological Review*, 81, 214-241.
- Swinney, D.A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 6, 645-659.

- Swinney, D.A., Onifer, W., Prather, P., & Hirshkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory and Cognition*, 7, 159–165.
- Tabossi, P. (1982). Sentential context and the interpretation of unambiguous words. *Quarterly Journal of Experimental Psychology*, 34A, 79–90.
- Tabossi, P. (1983). *Interpreting words in context*. D.Phil. Thesis, University of Sussex.
- Tabossi, P., & Johnson-Laird, P.N. (1980). Linguistic context and the priming of semantic information. *Quarterly Journal of Experimental Psychology*, 32, 595–603.
- Tanenhaus, M.K., & Lucas, M.M. (1987). Context effects in lexical priming. *Cognition*, 25, this issue.
- Templin, M.C. (1957). *Certain language skills in children: Their development and interrelationships*. Minneapolis: University of Minnesota Press.
- Tweedy, J.R., Lapinsky, R.H., & Schvaneveldt, R.W. (1977). Semantic context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory and Cognition*, 5, 84–98.
- Weinreich, U. (1966). Explorations in semantic theory. In T.A. Sebeok (Ed.) *Current Trends in Linguistics*, Vol. 3. The Hague: Mouton.
- Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.
- Wykes, T., & Johnson-Laird, P.N. (1977). How do children learn the meanings of verbs? *Nature*, 268, 326–327.
- Zwicky, A., & Sadock, J.M. (1973). Ambiguity tests and how to fail them. *Working Papers in Linguistics*. Ohio State University.

### Résumé

Cet article présente cinq phénomènes concernant le sens des mots. Il s'agit de (1) notre accès introspectif, limité quant à la nature des représentations lexicales; (2) l'existence d'entrées lexicales qui rendent accessible le sens d'un mot; (3) les effets du contexte sur l'interprétation des mots; (4) les lacunes systématiques dans l'acquisition du savoir lexical; et (5) l'existence de différents types sémantiques de mots appartenant à la classe ouverte. Ces phénomènes servent de point de départ à une théorie psychologique du sens des mots.