

Chapter 3

Do Word Meanings Exist?

This chapter addresses the question of whether word meanings exist in isolation at all. The proposal here is that, strictly speaking, words in isolation have meaning potential rather than meaning, and that actual meanings are best seen as events, only coming into existence when people use words, putting them together in clauses and texts. Corpus linguistics now enables us to study the contexts that activate different components of a word's meaning potential: to build an inventory of prototypical phraseology and relate meanings to the prototypes. One of the most encouraging discoveries of corpus linguistics has been that the prototypical phraseology associated with most words is highly patterned and that the number of patterns is manageably small. The undeniable creativity with which people use words is more than matched by—and dependent on—the patterned norms of usage for each word, with which particular meanings can be associated.

3.1 A Serious Question

Before we embark on a study of what words mean, and how meaning relates to use, it makes sense to ask whether words do in fact have meaning at all. The question is a serious one, and it is being asked by lexicographers, of all people. For example, Sue Atkins, a well-known lexicographer, is quoted by Kilgariff (1997) as saying, "I don't believe in word senses." This skepticism also relates to other theoretical aspects of language and thought. Nida, a pioneer of translation theory, argues (1997) that words have no meaning outside of context (linguistic or cultural).

There is an interesting and debatable relationship between this point of view and the principle of compositionality in predicate calculus, introduced by Frege (1884). The principle of compositionality regards words as the bricks used to build propositions. Just as a brick is not a building, so a word in itself is not meaningful. Frege argued that terms expressing concepts are variables that have meaning only when they are put together to form propositions. Only a proposition can be true or false. '*Snow is white*' may be true or false—the

conditions that determine whether a proposition is true or false can be stated explicitly. This principle was taken up by the young Wittgenstein and applied as a definition of language:

4.001 The totality of propositions is language. (Wittgenstein 1922)

Later it was developed by members of the Vienna Circle ('logical positivists'), as represented for example by Ayer (1936), who equated the meaning of a proposition with the conditions under which it can be verified. Thus, Frege's successors equated meaning with truth conditions of propositions. A consequence of this view is that a term in isolation—*snow*, for example—is meaningless.

As we saw in chapter 1, Wittgenstein himself subsequently (1953) moved away from this aspect of his earlier thinking and argued that philosophy should pay much closer attention to the ways in which words are actually used. We will return to an evaluation of Wittgenstein's relevance to lexical analysis in section 11.4.3.

Words are not as colorless and uniform as bricks. Nevertheless, the notion that words have no meaning is still prevalent, though not quite as prevalent as the converse, less sophisticated folk belief that each word has a meaning that can be precisely defined.

All this raises questions of fundamental importance to word-sense disambiguation and dictionary making. If word senses don't exist, then there isn't much point in trying to describe them in a dictionary, disambiguate them, or indeed do anything else with them. The very term *disambiguate* presupposes what Fillmore (1975) has characterized as a "checklist theory" of meaning.¹ In this chapter, I argue that words do have meaning (of a sort), but that—strictly speaking—when they are isolated from context, they have meaning potential rather than meaning as such. Meanings are contextually bound. The relationship between meaning in language and meaning in logic needs to be reexamined.

Vagueness and redundancy—features that are not readily compatible with a checklist theory—are important design features of natural language, which must be taken into account when doing serious natural-language processing. Words are so familiar to us, such an everyday feature of our existence, that it's an integral and prominent component of our psychological makeup, that it's hard to see what mysterious, complex, vague-yet-precise entities meanings are.

3.2 Common Sense

The notion that word meaning is mysterious may seem counterintuitive. To take a timeworn example, it seems obvious that the noun *bank* has at least two

senses: 'slope of land alongside a river' and 'financial institution'. But this line of argument is a honey trap. In the first place, these are not, in fact, two senses of a single word; they are two different words that happen to be spelled the same. They have different etymologies and different uses, and the only things they have in common are their spelling, pronunciation and inflection. Obviously, computational procedures for distinguishing homographs are both desirable and possible. But in practice they don't get us very far along the road to text understanding.

Linguists used to engage in the practice of inventing sentences such as '*I went to the bank*' and then claiming that they are ambiguous because they invoke both meanings of *bank* equally plausibly. Following Stock (1984), I have pointed out elsewhere (Hanks 1988) that in practice *bank* is rarely ambiguous, because terms in the immediate environment such as *money*, *clerk*, and *vaults* activate the 'financial institution' sense, while the 'riverside' sense is picked out by cooccurrence with terms such as *swim*, *canal*, and *water*. These collocations do not have to be in a structured syntactic relationship with the target word—they can be anywhere in the immediate environment.

Contextual clues disambiguate, and they can be computed to make choice possible, using statistical procedures such as those developed by Church and Hanks (1989). On the one hand, we find expressions such as these:

people without bank accounts; his bank balance; bank charges; gives written notice to the bank; in the event of a bank ceasing to conduct business; high levels of bank deposits; the bank's solvency; a bank's internal audit department; a bank loan; a bank manager; commercial banks; clearing banks; high-street banks; European and Japanese banks; a granny who tried to rob a bank (examples from the Cobuild corpus of 1983, cited in Hanks 1986b, 1987)

And on the other hand:

the grassy river bank; the northern bank of the Glen water; olive groves and sponge gardens on either bank; generations of farmers built flood banks to create arable land; many people were stranded as the river burst its banks; she slipped down the bank to the water's edge; the high banks towered on either side of us, covered in wild flowers

(examples from the Cobuild corpus of 1983, cited in Hanks 1986b, 1987)

Thus, the two words *bank* are not normally confusable in ordinary usage. In a random sample of 1,000 occurrences of the noun *bank* in the BNC, none were found where the 'riverside' sense and the 'financial institution' sense were both equally plausible. However, this merely masks the real problem,

bank, or an *ephemera bank*—and expect to be understood? The answer may well be yes, but either way, we need to know why.)

Another question: Is a bank₁ (financial institution) always an abstract entity? Then what about (2)?

(2) [He] assaulted them in a *bank* doorway.

Evidently the reference in (2) is to a building that houses a financial institution, not the institution itself. Do we want to say that the institution and the building that houses it are separate senses? Or do we go along with Pustejovsky, who would say that they are all part of the same “lexical conceptual paradigm (lep)” (Pustejovsky 1995), even though the superordinate semantic types [[Financial Institution]] and [[Building]] are different?

The lep provides a means of characterizing a lexical item as a meta-entry. This turns out to be very useful for capturing the systematic ambiguities which are so pervasive in language. ... Nouns such as *newspaper* appear in many semantically distinct contexts, able to function sometimes as an organization, a physical object, or the information contained in the articles within the newspaper.

- a. The newspapers attacked the President for raising taxes.
- b. Mary spilled coffee on the newspaper.
- c. John got angry at the newspaper.

(Pustejovsky 1995)

So it is with *bank*₁. Sometimes a bank is an institution; sometimes it is the building that houses the institution; sometimes it is the people within the institution who make the decisions and transact its business.

*Bank*₂ has similarly vague semantic boundaries. Does the ‘riverside’ sense always entail sloping land? Then what about (3)?

(3) A canoe nudged a *bank* of reeds.

Is a bank always beside water? Does it have one slope or two (one on either side)? Is it always dry land? Is it possible to reconcile (3), (4), and (5) as different aspects of the same meaning—different aspects of the same type of entity?

(4) Philip ran down the bracken *bank* to the gate.

(5) The eastern part of the spit is a long simple shingle *bank*.

3.3 Ockham’s Razor

Should the different uses of *bank*₂ in (3), (4), and (5) be treated as separate senses? Or should we apply Ockham’s razor, seeking to avoid a needless

which is vagueness: that is, in some cases *neither* of the meanings of *bank* just mentioned is fully activated, and yet communication succeeds perfectly satisfactorily.

(1) Averaging of this kind was first done at the National Hospital in London in the 1940s by George Dawson, initially using a technique of photographic superimposition and then later a system of addition using banks of condensers.

For example, in (1) it may not be immediately clear to the nonspecialist reader (nor even to specialists, perhaps) whether the *banks* of *condensers* are places where condensers are put for safekeeping (like a blood bank and a sperm bank) or whether they are sloping physical arrays (like a riverbank). Even if this question is decidable (which it may or may not be), devoting effort to deciding it smacks of pointless pedantry. Such a reader may visualize the banks of condensers mentioned in (1) as *either* a sloping array *or* a place where condensers are kept safely *or both*, without suffering major communication failure.

The obvious solution on encountering this sort of problem, you might think, would be to add more senses to the dictionary. And indeed this is often done. But it is not always a satisfactory solution, for a variety of reasons. For one, these doubtful cases (more examples are given below) do invoke one or another of the main senses to some extent, but only partially. Listing them as separate senses fails to capture the overlap and delicate interplay among them. It fails to capture the imprecision that is characteristic of words in use. And it fails to capture the dynamism of language in use. The problem is vagueness, not ambiguity. For the vast majority of words in use, including the two words spelled *bank*, one meaning can shade into another, and indeed the word may be used in a perfectly natural but vague or even contradictory way. In any random corpus-based selection of citations, a number of delicate questions arise that seem quite difficult to resolve or indeed unresolvable.

For example, how are we to regard expressions such as *data bank*, *blood bank*, *seed bank*, and *sperm bank*? Are they to be treated as part of the ‘financial institution’ sense? Even though no finance is involved, the notion of storing something for safekeeping is central. Or are we to list these all as separate senses (or as separate lexical entries), depending on what is stored? Or are we to add a ‘catchall’ definition of the kind so beloved of lexicographers: ‘any of various other institutions for storing and safeguarding any of various other things’? (But isn’t that insufficiently constrained? What precisely is the scope of ‘any of various’? Is it just a lexicographer’s cop-out? Is a speaker entitled to invent any old expression—say, a *sausage bank*, or a *restaurant*

- CARRIES OUT TRANSACTIONS
- CONSISTS OF A STAFF OF PEOPLE

And that of *bank*₂ like this:

- IS LAND
- IS SLOPING
- IS LONG
- IS ELEVATED
- IS SITUATED BESIDE WATER

On any occasion when the word *bank* is used by a speaker or writer, he or she activates some combination of one of these two sets of components, but no one of them is a necessary condition for something being a *bank* in either or any of the word's senses. Are any of the components of *bank*₂ necessary?

"IS LAND"? But think of a bank of snow.

"IS SLOPING"? But think of a path that runs along the bank of a canal or river.

"IS LONG"? But think of the bank around a pond or small lake.

"IS ELEVATED"? But think of the banks of rivers in East Anglia, where the difference between the water level and the land may be almost imperceptible.

"IS SITUATED BESIDE WATER"? But think of a grassy bank beside a road or above a hill farm.

3.4 Peaceful Coexistence of Incompatible Components

Components of the meaning of nouns (at least) are probabilistic and prototypical: the word *typically* should be understood before each of them. A bank is typically land, typically sloping, typically long, typically elevated, typically situated beside water. Whenever the word *bank* is used, some combination of these components is activated, but not necessarily all of them. There are no necessary conditions for meaning, but some at least must be present if the word is to be used successfully to create a meaningful communicative event. It is an open question how many components constitute a quorum for effective communication. The relationship between the reality of language in use and the widespread assumption that there must be necessary conditions for word meaning is in process of reappraisal. Is this a legacy of Aristotle, or is it something more basic and instinctual, arising from a biological need to believe that our cognitive classifications are certain, and therefore a reliable basis for decisive action?

multiplicity of entities? How delicate do we want our sense distinctions to be? Are 'river bank', 'sand bank', and 'grassy bank' three different senses? Can a sand bank be equated with a shingle bank?

Then what about 'a bank of lights and speakers'? Is it yet another separate sense, or a metaphorical variation? If we regard it as a metaphorical variation, we run into the problem that it has a different superordinate semantic type—[[Furniture]], rather than [[Land]]. Does this matter?

There is no single correct answer to such questions. Rather, the answer is determined by the user's intended application, or it may even be a matter of taste. Because there are many possible ways to carve up the semantic space of a word's meaning, lexicographers sometimes classify themselves as 'lumpers' or 'splitters': on the one hand, those who prefer—or, as it may be, who are constrained by marketing considerations—to lump uses together in a single sense; and on the other hand, those who prefer or are constrained to isolate fine distinctions. It is possible to multiply abstract entities such as sense distinctions ad nauseam, and this is indeed the natural instinct of the lexicographer. As new citations are amassed, new definitions are added to the dictionary to account for those citations that do not fit the existing definitions. As more and more evidence accumulates, this creates a combinatorial explosion of analytic problems, while still leaving many actual uses unaccounted for. Less commonly asked is the question, "Should we adjust the wording of an existing definition, to give a more generalized meaning?" But even if we ask this question, it is often not obvious how it is to be answered within the normal structure of a set of dictionary definitions.

Is there then no hope? Is natural language terminally intractable? Probably not. Human beings seem to manage all right. Language is certainly vague and variable, but it is vague and variable in principled ways, which are at present imperfectly understood. Pustejovsky's lexical conceptual paradigm offers one source of comfort, though there are considerable difficulties in translating it into practical procedures for traditional lexicography.

One could start by identifying the main semantic components of *bank* as separate, combinable, exploitable entities. This turns out to reduce the number of separate dictionary senses dramatically. The meaning of *bank*₁ can be expressed like this:

- IS AN INSTITUTION
- IS A LARGE BUILDING
- FOR STORAGE
- FOR SAFEKEEPING OF THINGS
- ESP. FOR SAFEKEEPING OF FINANCE/MONEY

Meaning components do not even have to be mutually compatible. The notion that something is at one and the same time an institution and a building, for example, may seem incompatible or incoherent, but that only means that these two components are never activated simultaneously (except in puns—see section 8.4—and certain other kinds of word play). They can still coexist peacefully as part of the word's meaning potential. By taking different combinations of components and showing how they combine, we can account economically and satisfactorily for the meaning of a remarkably large number of natural, ordinary uses.

This probabilistic componential approach also allows for vagueness.

(6) Adam sat on the bank among the bulrushes.

Is the component IS SLOPING present or absent in (6)? The question is irrelevant: the component is potentially present, but not active. But it is possible to imagine continuations in which it suddenly becomes very active and highly relevant—for example, if Adam slips *down* the bank and into the water.

If our analytic pump is primed with a set of probabilistic components of this kind, other procedures can be invoked. For example, semantic inferences can be drawn from superordinates: IS A BUILDING implies HAS A DOORWAY (see (1)); IS AN INSTITUTION implies HAS COGNITIVE MEMBERS (see (7)).

(7) The bank defended the terms of the agreement.

What's the downside? Well, it's not always clear which components are activated by which contexts. Against this: if it's not clear to a human being, then it can't be clear to a computer: whereas if it's clear to a human being, then in principle it should be possible to state the criteria explicitly and compute over them. A new kind of phraseological dictionary is called for, showing how different aspects of word meaning are activated in different contexts, and what those contexts are, taking account of vagueness and variability in a precise way.

A corpus-analytic procedure for counting how many times each feature is activated in a collection of texts will have considerable predictive power. After examining even quite a modest number of corpus lines, we naturally begin to form hypotheses about the relative importance of the various semantic components to the normal uses of the word, and how they normally combine. In this way, a default interpretation can be calculated for each word, along with a range of possible variations.

3.5 Meaning Events and Meaning Potentials

What, then, is a word meaning? It is reasonable to assume that in the everyday use of language, meanings are events, not entities (see Hanks 1994). Do meanings also exist outside the transactional contexts in which they are used? I would argue that they do not. It is a convenient shorthand to talk about "the meanings of words in a dictionary," but strictly speaking these are not meanings at all. Rather, they are **meaning potentials**—potential contributions to the meanings of texts and conversations in which the words are used, and activated by the speaker who uses them. This, and the fact that a word's meaning potential may contain incompatible components, suggests that meanings are best viewed as dynamic events that involve cooperation between speakers and hearers or writers and readers, rather than static, finitely bounded abstract objects. More will be said about meaning potentials in chapter 4.

The richness and complexity of a word's meaning potential vary considerably from context to context. Some terms denoting classes of physical objects have comparatively little potential for contextual differentiation. With all due respect to Atkins and Kilgarriff (as quoted above), it is obvious to everyone that an *elephant* is different from a *toothpick*, and it is possible to say quite a lot about the meaning of both terms, with reasonable confidence, without requiring contextualization. In other words, the default implicature of both terms is highly constrained. That is how many people believe that the meaning of all words functions. On the other hand, the boundaries of what counts as a toothpick (and even an elephant) are fuzzy. Traditional semanticists would worry about necessary conditions for being a toothpick: Is it still a toothpick even if it is never used for picking teeth? Suppose an object was intended to be a toothpick by its maker, but no one recognizes it as such? Does a toothpick have to be manufactured? Suppose a chimpanzee picks up a splinter of wood randomly and uses it to pick his teeth: is that splinter then a toothpick? How big can a toothpick be before it ceases to be a toothpick and must be called something else? Is *giant toothpick* a contradiction in terms?

A classic study of such dilemmas about the boundaries of word meaning is Labov (1973), in which Labov showed that people use different terms (*cup*, *vase*, or *bowl*) to name an identical outline shape, depending on the context—for example, whether it is used as a container for hot coffee, soup, rice, or flowers. However, attempts by Labov and his followers to lend mathematical precision to the degree of variability in each case have not won general acceptance.

with conventional implicatures, rather than by virtue of some underlying logic. The contribution of logical form (as opposed to valency) to meaning was overstated during the twentieth century.

3.6 Clause Structure and Wider Context

For most sentences, the meaning of a verb can be inferred with reasonable confidence from the semantic types of other words in the clause, provided that their semantic types are assigned. But in some cases, ambiguity remains. Let us now look at an extreme case where the meaning cannot be determined with confidence from the semantic values of the words in the clause structure, that is, from the phraseology of the immediate context.

The examples that follow are taken from various unpublished corpus-based lexicographical studies of the verb *check* during the 1980s and 1990s, as part of my work as a lexicographer. The purpose of this section is not to deny the contribution of immediate context to the interpretation of utterances; rather, it is to insist that, on the basis of immediate context and nothing else, meaning can be determined only probabilistically, not with certainty, and that this is more of a problem for corpus linguists, who take sentences or other little fragments of text from the middle of a discourse, than for ordinary readers and hearers, who are aware of the whole text from its inception, together with the domain and context of utterance (sports page in a newspaper, conversation about going shopping, or whatever) and who have already been primed by these contextual factors to form an opinion about the utterer's purpose and what sorts of events and participants are likely to be talked about in the discourse and therefore what the words used are likely to mean.

Check is a word of considerable syntactic complexity. Disregarding, for current purposes, *check*'s noun senses, the adjectival homograph denoting a kind of pattern (e.g., *a check shirt*), and various other distractions, we can zero in on the verb *check*. When this verb is transitive, its meaning potential, outside the domain of chess in which the word originated, has two different default interpretations: 'inspect' and 'cause to slow down or stop'. In the case of the second interpretation, as with many verbs in English, a transitive causative sense alternates with an intransitive inchoative: *a force can check an enemy's advance*, in which case *the enemy advances (and the enemy force) checks*.

It seems obvious that as a transitive verb, *check* cannot mean both 'inspect' and 'cause to slow down or stop' at the same time. In (8) and (9), these are obviously quite different meanings.

Another aspect of the variability of word meaning is illustrated by the word *oasis*. In English documents an oasis may, with approximately equal probability, be an area around a waterhole in a desert or a calm area in a busy city (or, with declining probability, any of various other things): context enables us to decide the relevant meaning. Ultimately, however, we should always bear in mind that choosing an interpretation is a matter for the reader or hearer. Corpus linguists sometimes speak as if interpretations spring fully fledged, untouched by human hand, from the corpus. They don't. The corpus contains traces of meaning events; the dictionary contains lists of meaning potentials. Mapping the one onto the other is a complex task, for which adequate tools and procedures remain to be devised.

However, the fact that the analytic task is complex does not necessarily imply that the results need to be complex. We may well find that the components of meaning themselves are very simple and that the complexity lies in establishing just how the different components combine.

In most ordinary sentences, verbs and nouns perform quite different clause roles. There is no reason to assume that the same kind of template or analytic procedure is appropriate to both. The relationship is like that between plug and socket. On the one hand, verbs assign contextual roles to the noun phrases in their environment. On the other hand, nouns (those eager suitors of verbs) have meaning potentials, activated when they fit (more or less well) into the verb frames. Together, they make meaningful language use possible. One of their functions, though not the only one, is to form propositions.

Propositions, not words, have entailments. But words can be used as convenient mental storage locations for conventional phraseology and for the entailments, implications, or implicatures that are associated with those bits of phraseology. **Implicature** is a term that was coined by H. P. Grice in 1968 to denote the act of intentionally implying a meaning that can be inferred from an utterance in context, but is neither explicitly expressed nor logically entailed by the statement itself. A famous example is the utterance, "Can you shut the door?" Although this is phrased as a question for the sake of politeness, it is in fact a request (or a command). Implicatures are like entailments, but weaker and probabilistic rather than logically necessary. An implicature is an act in which a speaker makes or relies on an implication, without providing sufficient evidence for hearers to judge that it is an entailment.

Implicatures play an important role in this book (see especially sections 4.2 and 11.6.1). TNE is a performance-driven theory of language: it argues that with very few exceptions, meaningful, communicative utterances in a natural language have meaning by virtue of their relation to conventional utterances

(8) It is not possible to **check** the accuracy of the figures.

(9) The DPK said that Kurdish guerrillas had **checked** the advance of government troops north of Sulaimaniya.

But then we come to sentences such as (10).

(10) The Parliamentary Assembly and the Economic and Social Committee were primarily or wholly advisory in nature, with very little **checking** power.

In (10), the meaning seems clear: the institutions mentioned had very little power either to inspect things or to stop things happening, or both. Perhaps an expert on European bureaucracy might be able to say whether one component or the other of *check* was more active, but the ordinary reader cannot be expected to make this choice, and the wider context is no help. Ordinary readers will get the general gist and move on without agonizing. Conveying meaning without forcing the reader to agonize is one of the design features of natural language. Only analysts agonize.

Arguably, therefore, the two meaning components of *check*, apparently in competition, coexist in (10) in a single use, as indeed they do in the noun phrase *checks and balances* and other similar uses. By relying too heavily on strongly contrastive examples such as (8) and (9), dictionaries have set up a false dichotomy, which extends to some but by no means all uses of a word. This tendency to polarize meanings for the sake of clarity of explanations is pervasive in dictionaries. Inconvenient borderline cases such as (10) tend to be swept under the carpet by lexicographers.

Perhaps word meaning exists at a more general level than is implied by the apparent precision and specificity of traditional dictionary definitions. The level of precision varies from context to context. The meaning of *checking* in (10) is quite general, but in other contexts very delicate implicatures can be deduced from the context. (11) and (12) are examples.

(11) Make sure you **check** the water level frequently.

(12) **Check** your telephone **directory** for the nearest branch.

Both (11) and (12) are clearly the 'inspect' sense, but there are subtle differences in implicature. Let us pause for a moment and explore these. In (11), there is a default implicature that the purpose of checking is to see if it is necessary to do something about the water level. In (12), on the other hand, the implicature is that the telephone directory is to be used as a source of information, not as something that needs to be corrected. Ordinarily, people check a telephone directory in order to obtain a telephone number or an address, not to make a change in the object inspected.

In some contexts, the meaning of *check* can be assigned neatly to one or other of these two implicatures. For example, *check the guttering* goes with (11); *check the map* goes with (12). Thus, it looks as if the default implicature—the underlying meaning—of *check* in (11) is different from its default implicature in (12). However, making such a distinction would, I think, be taking word meaning distinction too far—to an unreliably fine-grained level. To do so involves attributing too much to the immediate context and too little to the wider context, including the pragmatic context—the context of utterance.

In the first place, with regard to (12), no doubt there are people—proofreaders—whose job is to check telephone directories in the sense of ensuring that they are correct, although this is not the normal (or default) implicature of *check the telephone directory*. Proofreaders *check for errors*, not *for the nearest branch*. The verb *check* quite often takes a prepositional phrase governed by the preposition *for* in a purposive sense, but the purpose—what you are checking *for*—varies widely from context to context.² In the second place, there is no prepositional phrase in (11), and it is perfectly possible to imagine contexts in which you check the water level *for* information—in a water gauge, for example.

In other contexts, such as (13), the default implicature of the immediate context is undeniably ambiguous.

(13) the commander ... was **checking** the aircraft's instruments

If the event reported in (13) took place on the ground before takeoff, the commander was probably inspecting the instruments for errors—that is, to ensure they were working correctly. On the other hand, if the event took place during flight, it is much more likely that the commander was seeking information.

Similar uncertainty is found with less fine-grained semantic distinctions, too.

(14) On March 2nd 1872, a letter came from the Mayor saying that the Corporation's building would be ready in a few days and that they would be willing to receive patients under 'such arrangements as may be deemed best adapted to **check** the progress of the epidemic'.

In (14), the mayor surely wanted to stop the progress of the epidemic, not just to have a look at it. But selecting this sense of *check* is a pragmatic inference from the wider context and our knowledge of people's attitude to epidemics, not a necessary meaning condition attached to *check progress*. (15) illustrates the fact that *check the progress of something* can also have the 'inspect' sense,

while an invented contrastive example, (16), shows that precisely the same phrase can be used to activate the 'cause to slow down or stop' sense.

- (15) It was agreed that the situation would be reviewed in six months to check the progress of political reforms.
- (16) *It was agreed that action would be taken to check the progress of political reforms.*

I suppose that (16) could be uttered by a conservative who is opposed to political reforms. All this suggests that although it is often possible to draw delicate inferences from distinctions such as that exemplified in (11) and (12), such inferences must be held only lightly, tentatively, and provisionally. Word meaning typically operates at a very general level; a reader should draw inferences from text lightly, always being ready to revise them, while a lexical analyst should be wary of getting sucked into minutiae.

There are other cases where the wider context can enable a reader to decide the meaning of a text with greater confidence. Consider (17).

- (17) Then the boat began to slow down. She saw that the man who owned it was hanging on to the side and **checking** it each time it swung.
- Was the man in (17) inspecting it or was he stopping it? What is *it*? The boat, the side, or something else? The difficulty is resolved by looking back through the story leading up to this sentence—looking all the way back, in fact, to the first mention of *boat* in the document.

- (18) "Work it out for yourself," she said, and then turned and ran. She heard him call after her and got into one of the **swing boats** with a pale, freckled little boy ...

With this earlier context, it becomes clear that the boat in this story has nothing to do with vessels on water; it is a swinging ride at a fairground. The man in question (the owner), it turns out, is trying to slow down the swinging fairground boat (*checking it*) because a child in it is frightened. The reference of the term *boat* is established for the duration of the discourse and is obvious to readers who read the text from the beginning, but not to corpus linguists looking at a concordance.

This is a case where the relevant contextual clues are conclusive, though not in the immediate context. When, as corpus analysts, we plunge briefly *in medias res* and read a sentence or other fragment containing the word *boat*, we must not assume that *boat* is being used in its default meaning. To understand what the man was trying to do when he was checking the boat, we have to know that the scene is a fairground and the frame is a swing-boat ride, as

the wider context makes clear. If, as corpus analysts, we pay proper attention to textual cohesion, we are less likely to perceive ambiguity where there is none. Invented examples have no cohesion with anything.

Clues about particular or unusual word meanings that permeate the whole of a document are often to be found in its opening paragraphs. It is therefore desirable, in building a corpus, to select whole documents. Unfortunately, some corpora (including the BNC), in pursuit of the notion of a 'balanced' selection of texts, take a middle section of a text only, thus depriving the analyst of an important source of evidence.

A more complex example of the contribution of wider context to meaning, this time context following the target word, is (19).

- (19) Corporals checked kitbags and wooden crates and boxes ...

What were the corporals in (19) doing? It sounds at first as if they were inspecting something. But when we look at the continuation, the picture changes.

- (20) Sergeants rapped out indecipherable commands, corporals **checked** kitbags and wooden crates and boxes into the luggage vans.

In this case, the preposition *into* activates a causative and resultative sense of the verb, coercing the normal meaning of *check* away from the standard semantic component 'inspect' toward a meaning that is not normally associated with *check*, but rather with *put*. But this is not satisfactory either, because in its normal uses *put* is never causative. The verb *check* in the 'inspect' sense is not normally causative either, but in (20) it has been coerced by the context—in particular, the preposition *into*—into having the meaning 'cause to be put while inspecting and recording'. It seems unlikely that anyone will dispute that the meaning of (20) can be expressed as A (or, more broadly, B). This is remarkable, because this meaning cannot be derived compositionally from the meanings of the words actually used.

- A. [[Human]] cause [[Container]] to be **put** {into [[Vehicle]]}
- B. [[Human]] cause [[Physical Object]] to be **put** [Adv[Location]]

Is (20) a normal use of English? At first reading, it seems that the answer must be yes. It seems perfectly idiomatic and comprehensible. On the other hand, it does not precisely match any of the senses recorded in a large dictionary, and similar examples are hard to find. The *ODE* records a particle verb *check (something) in*, meaning 'to cause one's baggage to be weighed and put aside for consignment to the hold of an aircraft on which one is booked to travel'; also an intransitive prepositional verb *check into* (a hotel). But there is no standard transitive prepositional verb *check (something) into (something else)*. One way

of accounting for (20) is that it is an entirely unremarkable exploitation of the more standard expression *check (baggage)* in.

3.7 Where Corpus Analysis Runs Out

Before we leave *check* and its ambiguities, consider (21).

- (21) He soon returned to the Western Desert, where, between May and September, he was involved in desperate rearguard actions—the battle of Gazala, followed by Alamein in July, when Auchinleck checked Rommel, who was then within striking distance of Alexandria.

Without encyclopedic world knowledge, the fragment ... *Alamein in July; when Auchinleck checked Rommel* is profoundly ambiguous. It could mean either 'inspect' or 'cause to stop or slow down'.

I prepared an edited version of this sentence, (21'), omitting the words that clearly suggest a military context, and showed it to a group of English teenagers to whom the names *Auchinleck* and *Rommel* were unfamiliar.

- (21') *He soon returned to the desert, where, between May and September, he was involved in actions when Auchinleck checked Rommel.*

I asked the teenagers what *checked* means in this context. They were baffled, individually and collectively. The verb *check* was in their active vocabulary and some of them were chess players, but somehow the verb did not seem to fit. They agreed among themselves that probably, Auchinleck and Rommel were human beings. Was Auchinleck checking Rommel for fleas or for smuggled goods, one of them asked? Common sense may tell us that both interpretations are unlikely, but what textual clues are there to support common sense? Where does the assignment of meaning come from?

• From internal text evidence—in particular, the collocates? Relevant are the collocates *battle*, *rearguard actions*, and perhaps *striking distance*. Note that, although these collocates occur within the same text, they do not form part of the verb's valency preferences. They hardly seem close enough in the co-text to be conclusive; moreover, it is easy enough to construct counterexamples in the context of the same collocates, with a different meaning (e.g., *Before the battle, Auchinleck checked the deployment of his troops*).

• From the domain? If this citation were from a military history textbook, that might be a helpful clue. As it happens, the extract actually comes from an obituary in the *Daily Telegraph*, which the BNC very sensibly does not attempt to subelassify. And anyway, domain is only a weak clue. Lesk (1986) proposed

a disambiguation procedure based on the observation that the sorts of texts (including dictionary definitions) that talk about pine cones rarely also talk about ice-cream cones. This theme was elaborated in the claim by Yarowsky that there is generally only one sense of each polysemous word per discourse (see Gale, Church, and Yarowsky 1992).³ Be that as it may, in this case domain classification (as opposed to text differentiation) is unlikely to produce the desired result, since military texts do talk about both *checking equipment* and *checking the enemy's advance*.

• From real-world knowledge? Auchinleck and Rommel were generals on opposing sides. The name of a general may be used metonymically for the army that he commands, and real-world knowledge corroborates corpus evidence that armies *check* each other in the sense of halting an advance. This is probably close to psychological reality, but if this is all we have to go on, the difficulties of computing real-world knowledge satisfactorily start to seem insuperable.

• By assigning the terms *Auchinleck* and *Rommel* to some such lexical set as [[Human Military Role]]? This is similarly promising, but it relies on the existence of a metonymic exploitation rule of the following form:

[[Human = Military Commander]₁] checked [[Human = Military Commander]₂] = [[Army]₁] checked (i.e., halted the advance of) [[Army]₂]

This leads to the uncomfortable but inevitable conclusion that what seems perfectly obvious to a human being is deeply ambiguous to the more literal-minded computer and that there is no simpler way of resolving the ambiguity than by applying a rather complex rule of metonymy.

What's remarkable is that in the overwhelming majority of cases, a correct meaning can be assigned to a keyword on the basis of clues in its immediate environment—and in particular to a verb on the basis of the semantic values of its valencies. It is this simple fact that makes concordance-based corpus analysis possible at all. If language in use were less patterned—that is, if it were as 'creative' as some theoretical linguists have predicted, it would not be possible to tease out prototypical patterns of meaning and use from a concordance. Much of the rest of this book is devoted to showing how local contexts can be used to assign meanings to words in context.

3.8 Implications

The answer given in this chapter to the question asked at the beginning is that words in isolation are neither completely meaningful nor completely

meaningless; rather, they consist of clusters of semantic components constituting a meaning potential. For some but not all words, just one component of the meaning potential is cognitively salient: this is generally what people are referring to when they talk about a word's 'meaning'. Words that have just one salient meaning component exhibit what Sinclair (1991) called "the terminological tendency." An example is *microscope*, a term denoting a piece of scientific equipment with lenses for producing a magnified image of very small objects. That is its salient meaning component, but it has other, figurative meanings too.

The components of a word's meaning potential do not all have to be mutually compatible, as different combinations are activated in different contexts. So, yes, word meanings do exist, but traditional descriptions are misleading. Outside the context of a meaning event, in which there is participation of utterer and audience, words have meaning potential, not deterministic meaning. The meaning potential of each word in a text is activated cognitively in the mind of the reader or hearer by other words in the contexts in which it is used. These cognitive components are linked in a network that is both cognitive for each individual member of a language community and social for the language community as a whole. This network provides the whole semantic base of the language, with enormous potential for saying new things and relating the unknown to the known. The meaning potential of each lexical item in the network is in principle vague and variable—though some lexical items are more variable than others. Multiplicity of definitions is not necessarily a good way to go about accounting for the slight differences of meaning that are activated in different contexts.

Context and phraseology contribute substantially to the interpretation of words in text, but, as we saw in the case of the verb *check*, context and phraseology do not constitute a magic bullet for resolving all problems of ambiguity—only some of them.

A major future task for computational lexicography will be to identify meaning components, the ways in which they combine, relations with the meaning components of semantically related words, and the phraseological circumstances in which they are activated to make meanings.

3.9 Summary

- In isolation, words have **meaning potential**—the potential to make meanings when used.
- A meaning is an event; a meaning potential is a state.

- Words have meanings when they are used in a context (a verbal or situational context). Meanings are determined by the interaction of contexts and meaning potentials.
- A word's meaning potential consists of several semantic components, different combinations of which are activated in different contexts.
- For many but not all words, just one component of the meaning potential is salient; informally, this is often what is meant when people talk about a word's 'meaning'.
- Some semantic components of a word's meaning potential may be mutually incompatible.
- Mutually incompatible components are rarely activated simultaneously, except in puns and other word plays.