

LEXICAL
SEMANTICS

LEXEME

SENSE

The topics presented in this chapter serve to illustrate how much is missed by this narrow view. As we will see, the lexicon has a highly systematic structure that governs what words can mean, and how they can be used. This structure consists of relations among words and their meanings, as well as the internal structure of individual words. The study of this systematic, meaning related, structure is called **Lexical Semantics**.

Before moving on, we will first introduce a few new terms, since the ones we have been using thus far are entirely too vague. In particular, the word *word* has by now been used in so many different ways that it will prove difficult to make unambiguous use of it in this chapter. Instead, we will focus on the notion of a **lexeme**, an individual entry in the lexicon. A lexeme should be thought of as a pairing of a particular orthographic and phonological form with some form of symbolic meaning representation. The **lexicon** is therefore a finite list made up of lexemes. When appropriate, we will use the terms orthographic form, and phonological form, to refer to the appropriate form part of this pairing, and the term **sense** to refer to a lexeme's meaning component. Note that these definitions will undergo a number of refinements as needed in later sections.

Given this minimal nomenclature, let us return to the topic of what facts we can discover about lexemes that are relevant to the topic of meaning. A fruitful place to start such an exploration is a dictionary. Dictionaries are, after all, nothing if not repositories of information about the meanings of lexemes. Within dictionaries, it turns out that the most interesting place to look first is at the definitions of lexemes that no one ever actually looks up. For example, consider the following fragments from the definitions of *right*, *left*, *red*, *blood* from the *American Heritage Dictionary*(Morris, 1985).

right *adj* located nearer the right hand esp. being on the right when facing the same direction as the observer.

left *adj* located nearer to this side of the body than the right.

red *n* the color of blood or a ruby.

blood *n* the red liquid that circulates in the heart, arteries and veins of animals.

The first thing to note about these definitions is the surprising amount of circularity in them. The definition of *right* makes two direct references to itself, while the entry for *left* contains an implicit self-reference in the phrase *this side of the body*, which presumably means the *left* side. The entries for *red* and *blood* avoid this kind of direct self-reference by instead referencing each other in their definitions. Such circularity is, of course, inherent in all dictionary definitions, these examples are just extreme cases. In the end, all

definitions are stated in terms of lexemes that are, in turn, defined in terms of other lexemes.

From a purely formal point of view, this inherent circularity is evidence that what dictionaries entries provide are not, in fact, definitions at all. They are simply descriptions of lexemes in terms of other lexemes, with the hope being that the user of the dictionary has sufficient grasp of these other terms to make the entry in question sensible. As is obvious with lexemes like *red* and *right*, this approach will fail without some ultimate grounding in the external world.

Fortunately, even with this limitation, there is still a wealth of semantic information contained in these kinds of definitions. For example, the above definitions make it clear that *right* and *left* are similar kinds of lexemes that stand in some kind of alternation, or opposition, to one another. Similarly, we can glean that *red* is a color, it can be applied to both *blood* and *rubies*, and that *blood* is a *liquid*. As we will see in this chapter, given a sufficiently large database of facts such as these, many applications are quite capable of performing sophisticated semantic tasks (even if they do not *really* know their right from their left.)

To summarize, we can capture quite a bit about the semantics of individual lexemes by analyzing and labeling their relations to other lexemes in various settings. We will, in particular, be interested in accounting for the similarities and differences among different lexemes in similar settings, and the nature of the relations among lexemes in a single setting. This latter topic will lead us to examine the idea that lexemes are not unanalyzable atomic symbols, but rather have an internal structure that governs their combinatoric possibilities. Later, in Section 16.4, we will take a closer look at the notion of creativity, or generativity, and the lexicon. There we will explore the notion that the lexicon should not be thought of as a finite listing, but rather as a creative generator of infinite meanings.

Before proceeding, we should note that the view of lexical semantics presented here is not oriented solely towards improving computational applications of the more restrictive “only sentences have meaning” variety. Rather, as we will see, it lends itself to a wide array of applications that involve the use of words, and that could be improved by some knowledge of their meanings.

16.1 RELATIONS AMONG LEXEMES AND THEIR SENSES

The section explores a variety of relations that hold among lexemes and among their senses. The list of relations presented here is by no means exhaustive; the emphasis is on those relations that have had significant computational implications. As we will see, the primary analytic tool we will use involves the systematic substitution of one lexeme for another in some setting. The results of such substitutions can reveal the presence or absence of a specific relationship between the substituted lexemes.

Homonymy

HOMONYMY

We begin this section with a discussion of **homonymy**, perhaps the simplest, and semantically least interesting, relation to hold between lexemes. Traditionally, homonymy is defined as a relation that holds *between words that have the same form with unrelated meanings*. The items taking part in such a relation are called **homonyms**. A classic example of homonymy is *bank* with its distinct financial institution and sloping mound meanings, as illustrated in the following WSJ examples.

HOMONYMS

- (16.1) Instead, a *bank* can hold the investments in a custodial account in the client's name.
- (16.2) But as agriculture burgeons on the east *bank*, the river will shrink even more.

Loosely following lexicographic tradition, we will denote this relationship by placing a superscript on the orthographic form of the word as in **bank¹** and **bank²**. This notation indicates that these are two separate lexemes, with distinct and unrelated meanings, that happen to share an orthographic form.

It will come as little surprise that any definition this simple will prove to be problematic and will need to be refined. In the following discussion, we will explore this definition by examining pairs of words that satisfy it, but which for a number of reasons seem to be marginal examples. We will begin by focusing solely on issues of form, returning later to the topic of meaning. Note that while this may seem like an odd choice given the topic of this chapter, these discussions will serve to introduce a number of important distinctions needed in later sections. In this discussion, we will be primarily concerned with how well our definition of homonymy assists us in identifying and characterizing those lexemes which will lead to ambiguity problems for various applications.

Returning to the *bank* example, the first thing to note is that **bank¹** and **bank²** are identical in both their orthographic *and* phonological forms. Of course, there are also pairs of lexemes with distinct meanings which do not share *both* forms. For example, pairs like *wood* and *would*, and *be* and *bee*, are pronounced the same but are spelled differently. Indeed, as we saw in Chapter 5, when pronunciation in context is taken into account, the situation is even worse. Recall, that the lexemes *knee*, *need*, *neat*, *new*, *you*, *the*, and *to* can all be pronounced as [ni], given the right context. Clearly, if the notion of form in our definition of homonymy includes a word's phonological form in context, there will be a huge number of homonyms in English.

Of course, none of these examples are traditionally considered good candidates for homonymy. The notion of homonymy is most closely associated with the field of lexicography, where normally only dictionary entries with identical **citation-forms** are considered candidates for homonymy. Citation-forms are the orthographic forms that are used to alphabetically index words in a dictionary, which in English correspond to what we have been calling the root form of a word. Under this view, words with the same pronunciation but different spellings are not considered homonyms, but rather **homophones**, distinct lexemes with a shared pronunciation.

CITATION-FORMS

HOMOPHONES

Of course, there are also pairs of lexemes with identical orthographic forms with different pronunciations. Consider, for example, the distinct fish and music meanings associated with the orthographic form *bass* in the following examples.

- (16.3) The expert angler from Dora, Mo., was fly-casting for bass rather than the traditional trout.
- (16.4) The curtain rises to the sound of angry dogs baying and ominous bass chords sounding.

While these examples more closely fit the traditional definition of homonymy, they would only rarely appear in any traditional list of homonyms. Instead, lexemes with the same orthographic form with unrelated meanings are called **homographs**.

HOMOGRAPHS

Finally, we should note that lexemes with different parts of speech are also typically not considered to be good candidates for homonymy. This restriction serves to rule out examples such as *would* and *wood*, on grounds other than their orthography. The basis for this restriction is two-fold: first as we saw when we discussed part-of-speech tagging, lexemes with such different parts of speech are easily distinguished based on their differing syntactic environments, and secondly lexical items can take on many distinct

forms based on their inflectional and derivational morphology, which is in turn largely based on part-of-speech.

To complicate matters, the issue of differing morphology can also occur with lexemes that have the same part-of-speech. Consider the lexemes *find* and *found* in their *locating* and *creating an institution* meanings, as illustrated in the following WSJ examples.

- (16.5) He has looked at 14 baseball and football stadiums and found that only one -- private Dodger Stadium -- brought more money into a city than it took out.
- (16.6) Culturally speaking, this city has increasingly displayed its determination to found the sort of institutions that attract the esteem of Eastern urbanites.

Here we have two lexemes with distinct root forms, *find* and *found*, that nevertheless share the morphological variant *found* as the past tense of the first, and the root of the second.

At this point, having raised all of these complexities, we might create a more refined definition for homonymy as two lexemes with unrelated meanings, the same part of speech, and identical orthographic and phonological forms in all possible morphological derivations. Under this definition, all homonyms would also be both homographs and homophones, with the converse not necessarily being the case. Under this new definition, most of the homographs and homophones presented earlier would be ruled out as homonyms.

Such definitional exercises, however, merely obscure our reason for raising the issue of homonymy in the first place; homonymy is of interest computationally to the extent that it leads an application into dealing with ambiguity. Whether or not a given pair of lexemes cause ambiguity to arise in an application is entirely dependent on the nature of the application. As we will see in the following discussion of various applications, distinguishing perfect examples of homonymy from imperfect examples is of very little practical value. The critical issue is whether the nature of the form overlap is likely to cause difficulties for a given application.

In **spelling correction**, homophones can lead to real-word spelling errors, or malapropisms, as when lexemes such as *weather* and *whether* are interchanged. Note that this is a case where a phonological overlap causes a problem for a purely text-based system. Additional problems in spelling correction are caused by such imperfect homographs as *find* and *found*, which have partially overlapping morphologies. In this case, a word-form like

founded may represent a correct use of the past tense, or an incorrect over-application of the regular past tense rule to an irregular verb.

In **speech recognition**, homophones such as *to*, *two* and *too* cause obvious problems. What is less clear, however, is that perfect homonyms such as *bank* are also problematic. Recall that speech recognition systems rely on language models that are often based on tables of N-gram probabilities. For perfect homonyms, the entries for all the distinct lexemes are conflated despite the fact that the different lexemes occur in different environments. This conflation results in inappropriately high probabilities to words that are cohorts of the lexeme not in use, and lower than appropriate probabilities to the correct cohorts.

Finally, **text-to-speech** systems are vulnerable to homographs with distinct pronunciations. This problem can be avoided to some extent with examples such as *conduct* whose different pronunciations are associated with the distinct parts of speech through the use of part-of-speech tagging. However, for other examples like *bass* the two lexemes must be distinguished by some other means. Note that this situation is the reverse of the one we had with spelling correction, here a fundamentally speech-oriented system is being plagued by an orthographic problem.

Polysemy

Having muddied the waters discussing issues of form and homonymy, let us return to the topic of what it means for two meanings to be related or unrelated. Recall that the definition of homonymy requires that the lexemes in question have distinct and unrelated meanings. This is the crux of the matter; if the meanings in question are related in some way then we are dealing with a single lexeme with more than one meaning, rather than two separate lexemes. This phenomenon of a single lexeme with multiple related meanings is known as **POLYSEMY**. Note that earlier we had defined a lexeme as a pairing between a surface form and a sense. Here we will expand that notion to be a pairing of a form with a set of related senses.

To make this notion more concrete, consider the following *bank* example from the WSJ corpus.

- (16.7) While some *banks* furnish sperm only to married women, others are much less restrictive.

Although this is clearly not a use of the sloping mound meaning of *bank*, it just as clearly is not a reference to a promotional giveaway at a financial institution. One way to deal with this use would be to create **bank³**, yet

another distinct lexeme associated with the form *bank*, and give it a meaning appropriate to this use. Unfortunately, according to our definition of homonymy, this would require us to say that the meaning of *bank* in this example is distinct and unrelated to the financial institution sense, which seems to be far too strong a statement. The notion of polysemy allows us to state that this sense of *bank* is related to, and possibly derived from, the financial institution sense, without asserting that it is a distinct lexeme.

ETYMOLOGY

As one might suspect, the task of distinguishing homonymy from polysemy is not quite as straightforward as we made it seem with these *bank* examples. There are two criteria that are typically invoked to determine whether or not the meanings of two lexemes are related or not: the history, or **etymology**, of the lexemes in question, and how the words are conceived of by native speakers. In practice, an ill-defined combination of evidence from these two sources is used to distinguish homonymous from polysemous lexical entries. In the case of *bank*, the etymology reveals that **bank¹** has an Italian origin, while **bank²** is of Scandinavian origin, thus encouraging us to list them as distinct lexemes. On the other hand, our belief that the use of *bank* in Example 16.7 is related to **bank¹** is based on introspection about the similarities of their meanings, and the lack of any etymological evidence for an independent third sense.

In the absence of detailed etymological evidence, a useful intuition to use in distinguishing homonymy from polysemy is the notion of coincidence. Cases of homonymy can usually be understood easily as accidents of history – two lexemes which have coincidentally come to share the same form. On the other hand, it is far more difficult to accept cases of polysemy as coincidences. Returning again to our *bank* example, it is difficult to accept the idea that the various uses of *bank* in all of its various repository senses are only coincidentally related to the savings institution sense.

Once we have determined that we are dealing with a polysemous lexeme, we are of course still left with the task of managing the potentially numerous polysemous senses associated with it. In particular, for any given *single* lexeme we would like to be able to answer the following questions.

- What distinct senses are there?
- How are these senses related?
- How can they be reliably distinguished?

The answers to these questions can have serious consequences for how semantic analyzers, search engines, generators, and machine translation systems perform their respective tasks. The first two questions will be covered

here and in Section 16.4, while the final question will be covered in depth in Chapter 17.

The issue of deciding how many distinct senses should be associated with a given polysemous lexeme is a task that has long vexed lexicographers, who until recently have been the only people engaged in the creation of large lexical databases. Most lexicographers take the approach of creating entries with as many senses as necessary to account for all the fine distinctions in meaning observed in some very large corpus of examples. This is a reasonable approach given that the primary use for a traditional dictionary is to assist users in learning the various uses of a word. Unfortunately, it tends to err on the side of making more distinctions than are normally required for any reasonable computational application.

To make this notion of distinguishing distinct senses more concrete, consider the following uses of the verb *serve* from the WSJ corpus.

- (16.8) They rarely *serve* red meat, preferring to prepare seafood, poultry or game birds.
- (16.9) He *served* as U.S. ambassador to Norway in 1976 and 1977.
- (16.10) He might have *served* his time, come out and led an upstanding life.

Reasonable arguments can be made that each of these examples represents a distinct sense of *serve*. For example, the implicit contrast between *serving red meat* and *preparing seafood* in the first example indicates a strong connection between this sense of *serve* and the related notion of food preparation. Since there is no similar component in any of the other examples, we can assume that this first use is distinct from the other two. Next, we might note that the second example has a different syntactic subcategorization from the others since its first argument, which denotes the role played by the subject, is a prepositional phrase. As will be discussed in Section 16.3, such differing syntactic behaviors are often symptomatic of differing underlying senses. Finally, the third example is specific to the domain of incarceration. This is clear since this example provides almost no specific information about prison, and yet has an obvious and clear meaning; a meaning which plays no role in the other examples.

Another practical technique, for determining if two distinct senses are present is to combine two separate uses of a lexeme into a single example using a conjunction, a device has the rather improbable name of **zeugma**. Consider the following ATIS examples.

ZEUGMA

- (16.11) Which of those flights serve breakfast?
- (16.12) Does Midwest express serve Philadelphia?

(16.13) ?Does Midwest express serve breakfast and Philadelphia?

The oddness of invented third example indicates there is no sensible way to make a single sense of *serve* work for both breakfast and Philadelphia. More precisely, the underlying concepts invoked by *serve* in the first example can not be applied in any meaningful way to *Philadelphia*. This is an instance where we can make use of examples from a corpus along with our native intuitions in a structured way to discover the presence or distinct senses.

WORD SENSE
DISAMBIGUA-
TION

The issue of discovering the proper set of senses for a given lexeme is distinct from the process of determining which sense of a lexeme is being used in a given example. This latter task is called **word sense disambiguation**, or **word sense tagging** by analogy to part-of-speech tagging, and is covered in detail in Chapter 17. As this analogy implies, the task typically presumes that a *fixed* set of senses can be associated with each lexical item, a dubious proposition that we will take up in Section 16.4.

Finally, let us turn briefly to the topic of relatedness among the various senses of a single polysemous lexeme. Earlier, we made an appeal to the intuition that the polysemous senses of a lexeme are unlikely to have come about by coincidence. This raises the obvious question that if they are not related by coincidence, how are they related. This question has not received much attention from those constructing large lexicons since as long as the lexicon contains the correct senses, how they came to be there is largely irrelevant. However, as soon as applications begin to deal with a wide variety of inputs, they encounter novel uses that do not correspond to any of the static senses in the system's lexicon. By examining the systematic relations among listed senses, we can gain insight into the meanings of such novel uses. These notions will be discussed in more detail in Section 16.4.

SYNONYMY

Synonymy

SUBSTI-
TUTABILITY

The phenomenon of synonymy is sufficiently widespread to account for the popularity of both thesauri and crossword puzzles. As with homonymy, the notion of **synonymy**, has a deceptively simple definition: *different lexemes with the same meaning*. Of course, this definition leaves open the question of what it means for two lexemes to mean the same thing. Although Section 16.3 will provide some answers to this question, we can make progress without answering it directly by invoking the notion of **substitutability**: two lexemes will be considered synonyms if they can substitute for one another in a sentence without changing either the meaning or the acceptability of the sentence. The following ATIS examples illustrate this notion of substi-

tutability.

- (16.14) How big is that plane?
- (16.15) Would I be flying on a large or small plane?

Exchanging *big* and *large* in these examples has no noticeable effect on either the meaning or acceptability of these sentences. We can take this as evidence for the synonymy of *big* and *large*, at least for these examples. Note that this is intended to be a very narrow statement. In particular, we are not saying anything about the relative likelihood of occurrence of *big* and *large* in contexts similar to these.

Not surprisingly, if we take the notion of substitutability to mean substitutable in all possible environments, then true synonyms in English are few and far between, as it is almost always possible to find some sentence where a purported synonym fails to substitute successfully. Given this, we will fall back on a weaker notion that allows us to call two lexemes synonyms if they are substitutable in *some* environment. This is, for all practical purposes, the notion of synonymy used in most dictionaries and thesauri.

The success or failure of the substitution of a given pair of candidate synonyms in a given setting depends primarily on four influences: polysemy, subtle shades of meaning, collocational constraints, and register. As we will see, only the first two involve the notion of meaning.

To explore the effect of polysemy on substitutability, consider the following WSJ example where a substitution of *large* for *big* clearly fails.

- (16.16) Miss Nelson, for instance, became a kind of big sister to Mrs. Van Tassel's son, Benjamin.
- (16.17) ?Miss Nelson, for instance, became a kind of large sister to Mrs. Van Tassel's son, Benjamin.

The source of this failure is the fact that the lexeme *big* has as one of its distinct polysemous senses the notion of being older, or grown up. Since the lexeme *large* lacks this sense among its many meanings, it is not substitutable for *big* in those environments where this sense is required. In this instance, the result is a sentence with a different meaning altogether. In other cases, such a substitution may result in a sentence that is either odd or entirely uninterpretable.

We referred to the next influence on synonymy as *shades of meaning*. By this, we have in mind cases where two lexemes share a central core meaning, but where additional ancillary facts are associated with one the lexemes. Consider the use of the lexemes *price* and *fare* in the ATIS corpus.

Semantically, both have the notion of the cost for a service at the core of their meanings. They are not, however, freely interchangeable. Consider the following ATIS examples.

- (16.18) What is the cheapest first class fare?
(16.19) ?What is the cheapest first class price?

Exchanging *price* for *fare* in this example leads to a certain amount of oddity. The source of this oddness is hard to pin down, but *fare* seems to be better suited to the costs for various services (ie. coach, business and first class fares), while *price* seems better applied to the tickets that represent these services. Of course, a more complete account of how these lexemes are used in this domain would require a systematic analysis of a corpus of examples. The point is that although these terms share a core meaning, there are subtle meaning-related differences that influence how they can be used.

These two influences on substitutability clearly involve the meanings of the lexical items. There are, however, other influences on the success or failure of a synonym substitution that are not based on meaning in any direct way. Collocational constraints are one such influence. By a collocational constraint, we mean the kind of arbitrary associations, or attractions, between lexical items that were captured using techniques such as N-grams in Chapter 6.

Consider the following WSJ example.

- (16.20) We frustrate 'em and frustrate 'em, and pretty soon they make a big mistake.
(16.21) ?We frustrate 'em and frustrate 'em, and pretty soon they make a large mistake.

As this example illustrates, there is a preference for using *big* rather than *large* when referring to mistakes of a critical or important nature . This is not due to a polysemy difference, nor does it seem to be due to any subtle shaded meaning difference between *big* and *large*. Note also, that this is clearly different than the *large sister* example in that a *large mistake* is still interpretable in the correct way; it just does not seem as natural to use *large* as *big*. Therefore, in this case, we must say that there is simply an arbitrary preference for *big* as opposed to *large* as applied to *mistakes*.

Finally, by **register**, we mean the social factors that surround the use of possible synonyms. Here we are referring to lexemes with essentially identical meanings that are not interchangeable in all environments due to factors such as politeness, group status, and other similar social pressures. For ex-

ample, multisyllabic lexemes with Latin or Greek origins are often used in place of shorter lexemes when a technical or academic style is desired.

As was the case with homonymy, these influences on synonymy have differing practical implications for computational applications. In Chapters 19 and 20, we will see that similarity of meaning, collocational constraints, and appropriateness of use are of great importance in natural language generation and machine translation. On the other hand, in the domains of information extraction and information retrieval, appropriateness of use is of far less consequence than the notion of identity of meaning.

Hyponymy

In our discussion of *price* and *fare*, we introduced the notion of pairs of lexemes with similar but non-identical meanings. The notion of **hyponymy** is based on a restricted class of such pairings: *pairings where one lexeme denotes a subclass of the other*. For example, the relationship between *car* and *vehicle* is one of hyponymy. Since this relation is not symmetric we will refer to the more specific lexeme as a **hyponym** of the more general one, and conversely to the more general term as a **hypernym** of the more specific one. We would therefore say that *car* is a hyponym of *vehicle*, and *vehicle* is hypernym of *car*.

HYPONYMY

HYPONYM

HYPERNYM

As with synonymy, we can explore the notion of hyponymy by making use of a restricted kind of substitution. Consider the following schema.

That is a *x*. \Rightarrow That is a *y*.

If *x* is a hyponym of *y*, then in any situation where the sentence on the left is true, the newly created sentence on the right must also be true, as in the following example.

That is a car. \Rightarrow That is a vehicle.

There are a number of important differences between this kind of limited substitution and the kind of substitutions discussed with respect to synonymy. There the resulting sentence could plausibly serve as a substitute for the original sentence. Here, the new sentence is not intended to be a substitution for the original, rather it is merely serves as diagnostic test for the presence of hyponymy.

The concept of hyponymy is closely related to a number of other notions that play central roles in biology, linguistic anthropology and computer science.

The term **ontology** usually refers to an analysis of some domain, or **microworld**, into a set of distinct objects. A **taxonomy** is a particular arrange-

ONTOLOGY

TAXONOMY

OBJECT HIERARCHY

ment of the elements of an ontology into a tree-like class inclusion structure. Normally, there are a set of well-formedness constraints on taxonomies that go beyond their component class inclusion relations. For example, the lexemes *hound*, *mutt*, and *puppy* are all hyponyms of *dog*, but it would be odd to construct a taxonomy from those pairs since the concepts motivating the relations is different in each case. Finally, the computer science notion of an **object hierarchy** is based the notion that objects from an ontology arranged in a taxonomy, can receive, or inherit, features from their ancestors in a taxonomy. This, of course, only makes sense when the elements in the taxonomy are in fact complex structured objects with features to be inherited.

Therefore, sets of hyponymy relations, by themselves, do not constitute an ontology, category structure, taxonomy, or object hierarchy. They have, however, proved to be useful as approximations to such structures. We will return to the topic of hyponymy in Section 16.2 when we discuss the WordNet database.

16.2 WORDNET: A DATABASE OF LEXICAL RELATIONS

The widespread use of lexical relations in linguistic, psycholinguistic, and computational research has led to a number of efforts to create large electronic databases of such relations. These efforts have, in general, followed one of two basic approaches: mining information from existing dictionaries and thesauri, and handcrafting a database from scratch. Despite the obvious advantages of reusing existing resources, WordNet, the most well-developed and widely used lexical database for English, was developed using the latter approach (Beckwith *et al.*, 1991).

WordNet consists of three separate databases, one each for nouns and verbs, and a third for adjectives and adverbs; closed class lexical items are not included in WordNet. Each of the three databases consists of a set of lexical entries corresponding to unique orthographic forms, accompanied by sets of senses associated with each form. Figure 16.1 gives some idea of the scope of the current, WordNet 1.6, release. The databases can be accessed directly with a browser (locally or over the Internet), or programmatically through the use of a set of C library functions.

In their most complete form, WordNet's sense entries consist of a set of synonyms, a dictionary-style definition, or gloss, and some example uses. Figure 16.2 shows an abbreviated version of the wordnet entry for the noun *bass*. As this entry illustrates, there are several important differences be-