PART I

Meaning and the lexicon
1 The lexicon – some preliminaries

Key words: lexical semantics, lexicon, lexis, mental lexicon, lexeme, conventional, (non-)compositional, arbitrary, lexical unit, word, idiom, collocation, open/closed class, lexicography, corpus, native speaker intuition

1.1 Overview

In order to set the stage for exploring lexical semantics, this chapter defines basic terms and ideas in the field, particularly the notions of lexicon, lexeme, and word. It then describes and evaluates four methods for investigating the lexicon: dictionaries, corpora, intuition, and experimentation.

1.2 What is a lexicon?

Word meanings are notoriously difficult to pin down – and this is well demonstrated in defining the basic terminology of lexical semantics. Semantics is the study of linguistic meaning, but it will take the next three chapters to discuss what meaning might mean in any particular theory of semantics. The lexical in lexical semantics refers to the lexicon, a collection of meaningful linguistic expressions from which more complex linguistic expressions are built. Such lexical expressions are often, but not always, words, and so lexical semantics is often loosely defined as ‘the study of word meaning,’ although the word word, as we shall see, is not the most straightforward term to use.

While many of the details of the structure and content of the lexicon are discussed in detail in later chapters, some general discussion of what the lexicon is and what it contains must come first. A lexicon is a collection of information about words and similar linguistic expressions in a language. But which information? Which expressions? What kind of collection? Whose collection? We’ll cover these issues in the following subsections, but first we must acknowledge the polysemy (the state of having multiple meanings) of the word lexicon. Lexicon can refer to:
4 lexical meaning

- a dictionary, especially a dictionary of a classical language; or
- the vocabulary of a language (also known as lexis); or
- a particular language user’s knowledge of her/his own vocabulary.

For our purposes, we can disregard the first meaning and leave the study of such dictionaries to students of classical languages. The last two definitions are both relevant to the study of lexical semantics. In speaking of the lexicon, different scholars and theories assume one or the other or the interrelation of both, as the next subsection discusses.

1.2.1 Where is the lexicon?

Some traditional approaches to the lexicon generally make claims about the vocabulary of a language, its lexis. Taking this perspective on vocabulary, the lexicon is “out there” in the language community – it is the collection of anything and everything that is used as a word or a set expression by the language community. Other linguistic perspectives, including those discussed in this book, focus on vocabulary “in here” – in the mind of a language user. The term mental lexicon is used in order to distinguish this more psychological and individualistic meaning of lexicon.

Clearly though, we have to take into account the fact that the “out there” and “in here” lexicons are interrelated; in order to communicate with each other, speakers of a language must aim to have reasonably similar ways of using and understanding the words they know – otherwise, if you said banana, I’d have no reason to believe that you didn’t mean ‘robot’ or ‘hallelujah.’ The lexicon of the language “out there” in our culture is the lexicon that we, as individuals, aim to acquire “in here” and use. This is not the same as saying that the lexicon of a language is a union of all the lexicons of all the language’s speakers. When linguists study a language’s lexicon, they tend to idealize or standardize it. For instance, say there’s an English speaker somewhere who, ever since being hit on the head with a mango, mistakenly uses the word goat to mean ‘pencil.’ Just because there’s someone who uses the language in this way does not mean that this fact about the use of the language needs to be accounted for in a model of the English lexicon – his use is clearly a mistake. So, in order to study the lexicon of a language, one needs to have a sense of what does and does not count as part of that language. Sometimes these decisions are somewhat arbitrary, but they are not simply decisions made on the basis of what is “correct” English in some school-teacherish (that is, prescriptive) sense. Non-standard words and uses of words are also part of the language that we want to explain, and we can pursue interesting questions by looking at them. For example, some people use wicked as slang for ‘especially good,’ which might lead us to ask: how is it that a word’s meaning might change so much that it is practically the opposite of what it originally meant?

Similarly, although mental lexicons exist in individual speakers’ minds, in studying the mental lexicon we do not necessarily want to investigate any one
particular speaker’s lexicon (otherwise, we would have another few billion lexicons to investigate after we finish the first one). Instead, the focus is usually on an imagined “ideal” speaker of the language, which again brings us back to the notion of a language’s lexicon. For an “ideal” mental lexicon, we imagine that a speaker has at her disposal the knowledge necessary to use the language’s lexis.

Most current approaches to the lexicon attempt to find a balance between the “out there” and the “in here.” While particular models of lexical meaning will be evaluated in this book on the basis of their psychological plausibility, part of what makes a theory psychologically plausible is whether it is consistent with (and engages with) the social facts of language acquisition and use. My continued use of the ambiguous term lexicon is an acknowledgment of the dual nature of the object of our study, but the terms mental lexicon and lexis are used wherever disambiguation is needed.

1.2.2 What’s in a lexicon?

Having discussed the where of the lexicon, we move on to the what. The things that one knows when one knows a language can be divided into two categories: the lexical and the grammatical. A grammar is a system of rules or regularities in a language, and a lexicon is (at the very least) a collection of linguistic knowledge that cannot be captured by rules. The grammar accounts for linguistic issues like word order and regular morphological and phonological processes. For instance, our grammar tells us the difference between the sentences Bears trouble bees and Bees trouble bears, and that this is the same kind of difference as the difference between The kangaroo ate a flower and A flower ate the kangaroo. What the grammar cannot tell us is what bear and bee and trouble bring to the sentence. At some point in our acquisition of English, we learned that the sound [bi] and the spelling b-e-e are paired with a particular set of linguistic and semantic properties – like being a noun and denoting a kind of insect. The lexicon is the collection of those associations between pronunciations, meanings, and grammatical properties that had to be learned rather than produced by grammatical rules.

The lexicon is organized into lexical entries, much as a dictionary is organized into entries that pull together all the information on a headword (the word, typically in boldface type, at the start of a dictionary entry). Each of these lexical entries collects the appropriate information about a particular linguistic expression, called a lexeme. (Later we look at why it is more precise to use the term lexeme rather than word in the study of lexical meaning.) In the remainder of this subsection, we consider which expressions are lexemes and belong in the lexicon, then in §1.2.3 we’ll consider what information goes in a lexeme’s lexical entry. Section 1.2.4 goes into more detail on the notion of a lexeme as an abstract representation and the relationship between a lexeme (and its lexical entry) and actual uses of an expression. Let’s start with this description of lexeme:
A linguistic form (i.e. a bit of speech and/or writing) represents a lexeme if that form is conventionally associated with a non-compositional meaning.

In order to make more sense of this, let’s look more closely at the concepts of conventionality and (non-)compositionality, in turn.

**Conventionality**

Lexemes, and the information about them in the lexicon, are conventional – that is, these form–meaning pairings are common knowledge among the speakers of the language, and we have had to learn these particular associations of form and meaning from other members of the language community. Compare, for example, a scream with a word. If you heard someone screaming, you would not know (without further information) what they were making noise about nor why they were making noise – after all, we scream with surprise, with delight, or with horror. But if the person yelled Spider! or Fire! or Jump!, you’d know what they were yelling about (but perhaps still not why they were yelling about it) because those words are used by members of our speech community to signal particular things.

**Non-compositionality**

Lexemes are non-compositional – that is, the meanings of these linguistic forms are not built out of (or predictable from) the meanings of their parts. For example, the word cat is non-compositional because its meaning is not evident from the sounds or letters that constitute the word. It’s not that the sound /t/ represents the tail of the cat or that the vowel /æ/ tells us that a cat has fur. The word cat and its meaning thus constitute an arbitrary pairing of form and meaning.

The meaning of black cat, however, is deducible if you know:

(a) the word classes (adjective, noun) and meanings of the words black and cat
(b) what it means in English to put an adjective in front of a noun.

Thus the meaning of the clause black cat is compositional; its meaning is built from the meanings of its parts. This means that black cat does not need to be included in the lexicon, but its non-compositional parts (black and cat) do.

**Lexemes and morphological complexity**

While lexical semantics is often loosely defined as ‘the study of word meaning,’ the use of word in this definition is misleading, since lexical semantics is more accurately described as the study of lexeme meaning, and (a) not all words are lexemes and (b) not all lexemes are words.

Here, we need to pause and define a little terminology from morphology, the study of word structure. Morphemes are the smallest meaningful units of language, so a word is one morpheme if it is not built out of smaller meaningful parts. So, for example, language is a single morpheme: if we try to divide it into
smaller parts (like *l* and *anguage* or *langu* and *age*), we don’t get meaningful parts of English. (And although we can see the word *age* in *language*, that is just accidental. *Age* is not a morpheme within *language* as ‘age’ is not part of the meaning of *language*.) On the other hand, *sentimentally* is composed of three morphemes: the noun *sentiment* and the suffixes *-al* (which is added to nouns to make adjectives and could be glossed as ‘marked by’) and *-ly* (which turns an adjective into an adverb: ‘in a certain manner’). This results in a complex word that means ‘in a manner that is marked by sentiment.’ *Sentimentally* is thus compositionally formed from *sentiment* plus two suffixes, and since it is compositional, in that the meaning comes straightforwardly from the combination of the parts, it does not count as a lexeme on our definition. However, *sentiment* and the suffixes *-al* and *-ly* act like lexemes, in that they have conventional meanings that cannot be derived from the meanings of their parts.

Suffixes, prefixes, and other bits of language that always attach to other morphemes are called **bound morphemes**: they must be bound to another linguistic form in order to be used. Words, on the other hand, are **free morphemes**, in that they can be used without being attached to anything else. Just as words can be arranged and rearranged to make new phrases with compositional meanings, morphemes can be arranged and rearranged to make new words with compositional meanings. The first time you come across a new morphologically complex word, like *unputdownable* or *pseudoscientifically*, you will be able to understand the word if you can understand its parts.

So far, we have seen that morphemes are lexemes in that they have conventional, non-compositional meanings. But more complex expressions may also be lexemes. For example, while *greenhouse* is a compound noun derived from two free morphemes, its meaning is not deducible from its parts, since a greenhouse is not actually green and it is debatable whether it is a house. A series of words can also be a single lexeme, as demonstrated by the bold expressions in the following examples:

(1) a. **Look up** that word in the dictionary!
    b. Look up! There’s a dictionary in that tree.

(2) a. Her beloved pet is now **pushing up daisies**.
    b. Fido was the **apple of her eye**.

In (1a), the **phrasal verb** *look up* is a single lexeme meaning ‘consult,’ as opposed to (1b), in which *look* and *up* each contributes its own meaning. **Idioms** like *pushing up daisies* (‘dead’) and *apple of one’s eye* (‘one’s beloved’) in (2) are also lexical expressions, since they are non-compositional, and thus have to be learned and mentally stored by a language speaker.

In summary, the term **lexeme** includes:

- **simple words** (free morphemes) that cannot be broken down into smaller meaningful parts, such as *cup*, *Cairo*, and *contribute*;
- **bound morphemes**, like *un-* as in *unhappy* and *-ism* as in *racism*;

© in this web service Cambridge University Press  www.cambridge.org
morphologically complex words whose meaning is not predictable from the meanings of the parts, including compounds like greenhouse (‘a glass building for growing plants in’) and needlepoint (‘a method of embroidery onto canvas’);

• set phrases whose meaning is not compositional, such as phrasal verbs like throw up (‘vomit’) and give up (‘quit’) and idioms like having the world on one’s shoulders and fly off the handle.

**Are all lexemes non-compositional?**

While linguists agree that a lexicon contains conventional, non-compositional form–meaning pairings, opinions differ as to whether one’s mental lexicon also contains some compositional expressions. In other words, just because we *could* understand the meaning of a complex expression from the meanings of its parts, doesn’t mean that we necessarily always go through the process of composing that complex expression every time we use it. In cases like those in (3), the expressions are so well-worn that they seem like idioms, in spite of having conventional meanings.

(3) a. happiness (happy + ness)
    b. How are you?
    c. It’s not whether you win or lose, it’s how you play the game.

Should the items in (3) be considered to be lexemes? One argument for including compositional expressions in the lexicon is that it is advantageous for a language user to have ready access to commonly used expressions. So, in a particular individual’s mind it might be convenient if complex and frequent but predictable expressions, like *How are you?* or *I love you*, were stored in the mental lexicon. Having such expressions available both through grammatical assembly and through ready-made lexemes is redundant – and thus it appears to require needless extra effort to learn and store the expression in the mental lexicon. But while non-redundant models of language are more streamlined and elegant, they are not necessarily more realistic. After all, representing the same expression (or access to the same expression) in different ways or different places in one’s mind could make language production and comprehension processes less likely to break down, since if one route to the expression fails, another might succeed.

Another argument for including compositional expressions in the lexicon is that some of them are particularly conventionalized – that is to say, people sometimes rely on “ready-made” compositional expressions instead of composing new ones. The extremes of such conventionalization are seen in compositional clichés like *in this day and age*, *cry like a baby*, or the example in (3c), but conventionalization of compositional expressions can be subtler too, as studies of collocations (particularly frequent word combinations) have shown. A case in point is example (4), which shows how the meaning ‘without milk added’ is indicated by different modifiers, depending on what is being modified.
The lexicon – some preliminaries

(4)  
- black coffee
- black tea – however, this can also mean non-herbal tea, so people often prefer to say tea without milk or tea with no milk
- plain chocolate, dark chocolate
- dry cereal

Logically speaking, there is no particular reason why milkless coffee is black and not plain, nor why dark chocolate is not called black chocolate (after all, it is about the same color as black coffee). As English speakers, we’ve just learned that these adjectives and nouns are best used in some combinations and not others. Similarly, some nouns go with particular verbs. For example, in English one asks (or poses) a question, but one poses a riddle (one does not ask it) and one makes a query (one neither asks nor poses it). Such facts lead some to argue that we should be thinking of the lexicon as including larger and looser combinations than just single words and set idiomatic phrases, or that we should see the lexicon as a network of syntactic (grammatical) and semantic relations among lexemes. While these are very interesting issues, the focus in this book is particularly on the non-compositional forms (mostly words) that must be in the lexicon, as the emphasis on non-compositional meanings is what makes lexical semantics distinct from other kinds of semantics. Information about collocations is still of interest to us, though, as it may be used to determine differences in particular words’ meanings.

1.2.3 What is in a lexical entry?

Within the lexicon, the collection of information pertaining to a lexeme is said to be its lexical entry, analogous to a dictionary entry. The information that must be stored about a lexeme is precisely that which is unpredictable, or arbitrary. At the very least, this means that we need to know the basics of the lexeme’s form and what meaning(s) it is associated with. When speaking of a word’s form we usually mean its pronunciation, but if we know it as a written word, then its spelling is part of its form, and if it is a part of a sign language, then its “pronunciation” is gestural rather than vocal. We only need to store in the lexicon the details of the lexeme’s form that are not predictable; so, for example, we do not need to store the facts that cat is made possessive by adding ’s or that the c in cat is usually pronounced with a slight puff of air – these facts are predictable by rules in the language’s grammar and phonology. As we shall see in the coming chapters, theories differ in what information about meaning is (or is not) included in the lexicon. For many modern theories, meaning is not in the lexicon, but is a part of general, conceptual knowledge (chapter 4). That is to say, the linguistic form is represented in the lexicon, but instead of its definition being in the lexicon as well, the lexical entry “points” to a range of concepts conventionally associated with that word. Other theories (e.g. in chapter 3) view the lexicon more like a dictionary, which provides basic definitions for words.
What other information is included in a lexical entry, again, differs from theory to theory. Most would say that a lexical entry includes some grammatical information about the word, for example its word class (or part of speech: noun, verb, etc.), and the grammatical requirements it places on the phrases it occurs in. For instance, the lexical entry for *dine* includes the information that it is a verb and that it is intransitive in that it takes no direct object, as shown in (5). *Devour*, on the other hand, is recorded as a verb that is transitive, so that it is grammatical with a direct object, but not without one, as shown in (6). Asterisks (∗) signal ungrammaticality.

(5) a. We dined.
   b. ∗We dined a big bowl of pasta.

(6) a. ∗We devoured.
   b. We devoured a big bowl of pasta.

We come back to these issues in chapter 7, where we consider whether the meanings of words might determine their word classes, or whether they are entirely arbitrary and need full specification in the lexicon.

We may also need information in the lexicon about which words go with which other words – for instance, the fact that *stark* collocates with *naked* but not with *nude* or the fact that the conventional antonym of *alive* is *dead* and not *expired*. We come back to some of these issues below and in chapter 6.

### 1.2.4 The abstract nature of lexemes

The last thing to say in this preliminary tour of the lexicon is that a lexeme is not the same as a word in real language use. Lexemes are, essentially, abstractions of actual words that occur in real language use. This is analogous to the case of phonemes in the study of phonology. A phoneme is an abstract representation of a linguistic sound, but the phone, which is what we actually say when we put that phoneme to use, has been subject to particular linguistic and physical processes and constraints. To take a psycholinguistic view, a phoneme is a bit of language in the mind, but a phone is a bit of language on one’s tongue or in one’s ear. So, the English phoneme /l/ is an abstract mental entity that can be realized in speech variously as, say, a “clear” [l] in *land*, or a “dark” [ɻ] in *calm*. The phoneme is, so to speak, the potential for those two phones.

Similarly, when we use a word in a sentence, it is not the lexeme in the sentence, but a particular instantiation (i.e. instance of use) of that lexeme. Those instantiations are called lexical units. Take, for example, the lexeme *cup*. It is associated with a range of meanings, so that we can use it to refer to:

- any drinking vessel, or
- a ceramic drinking vessel with a handle whose height is not out of proportion to its width, or
The lexicon – some preliminaries

(c) a hand shaped so that there is a depression in its upward facing palm, or
(d) the part of a brassiere that covers a breast.

But in a particular use, as in sentence (7), the lexical unit *cup* is assigned just one of those meanings – in this case, meaning (b).

(7) I prefer my coffee in cups, not mugs.

In (7), *cup* also occurs with a particular orthographic form and it is morphologically marked as a plural. So, in this case we can say that the lexical unit *cups* in (7) is an instantiation of the lexeme *cup*, which has the plural form, the standard spelling and the (b) meaning associated with that lexeme.

The lexical entry provides the information needed to use the lexeme as a lexical unit; that is, it sets the parameters of that lexeme’s potential. So, as speakers or writers, we select particular lexemes to be realized as lexical units in our utterances because the parameters set in the lexical entry are consistent with what we want to do with that word – for example, for it to have a particular meaning and fill the verb slot in a sentence. As listeners or readers we recognize the lexical unit as being an instantiation of a particular lexeme and interpret the meaning of the lexeme in ways that are consistent with the lexical entry and the context in which the lexical unit occurs.

1.3 What is a word?

1.3.1 Defining word

While the loose definition of *lexical semantics* is ‘the study of word meaning,’ the focus of this textbook is more precisely lexeme meaning. Nevertheless, as a matter of fact, most of the cases discussed in this book (and most of what is traditionally considered to be lexical semantics) involve lexemes that are words rather than bound morphemes or multi-word lexemes (such as idioms).

So, it is worthwhile to say a word about what words are. I have already hinted at one problem with using the term *word* as a technical term: *word* is ambiguous in that it could mean ‘lexeme’ as in (8) or ‘lexical unit’ as in (9):

(8) Colour and color are two spellings of the same word.

(9) There are eight words in the sentence *Jo’s pet ferret hates Jo’s mother’s pet ferret*.

But aside from the lexeme/lexical unit ambiguity of *word*, it can be tricky to determine which lexemes count as words in a particular language. Part of the reason for this is that the notion *word* can be defined in a number of ways, including: