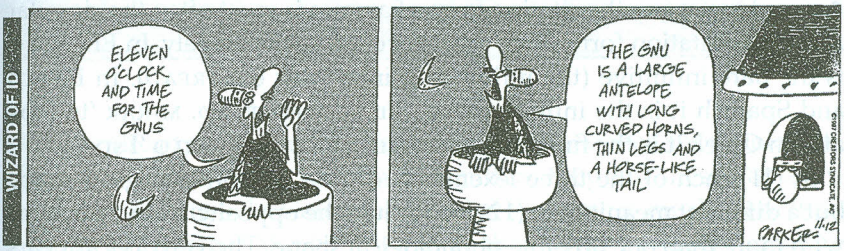
Content words	Function words
Nouns: *baby, bargain, Josianne*	Pronouns: *I, him, our*
Verbs: *publicize, hurtle, sleep*	Verbs: *am, was, should*
Adjectives: *peaceful, quick, bright*	Determiners: *the, an, a*
Adverbs: *readily, carefully*	Demonstratives: *this, those*
	Adverbs:[3] *very, not*
	Prepositions:[4] *in, by*

2.3.3 Lexemes

We need one more distinction, which will allow us to tell grammatical words apart from lexemes. Let's say we run across the word *dog* and notice that we have at least three tokens of the word:

(11) a. DOG$_1$: [noun], a canine
b. DOG$_2$: [noun], a hooked or U-shaped device used for gripping heavy objects
c. DOG$_3$: [verb], to follow closely and persistently



These three words sound alike: they are tokens of the same phonological word. Semantically, the meanings of (11a) and (11c) are related, and that of (11b) is not. Despite the semantic relatedness of (11a) and (11c), however, all three tokens can be said to have distinct meanings.

When we want to distinguish among phonologically similar forms on the basis of their differing meanings, as in (11), we call each a separate **lexeme**. A lexeme is a word with a specific sound and a specific meaning. Its shape may vary depending on syntactic context. Thus we

have *dog* and *dogs*, distinct grammatical word forms of the same lexeme DOG$_1$. (For more on the definition of lexeme, see the box opposite.) We use the term **paradigm** to refer to the set of all the inflected forms that a lexeme assumes.

A lexeme is a theoretical construct. It is not a sound form (e.g., *dog*), but rather a sign or set of signs, with sound form, syntax, and meaning all bound together. Because it stands outside any syntactic context beyond the one for which it is lexically specified or subcategorized, it is inherently unspecified for categories that are determined by context and expressed through inflection. Some linguists restrict the class of lexemes to the major lexical categories: noun, verb, and adjective/adverb (Aronoff 1994: 9–10).

DOG$_1$ 'a canine' and DOG$_2$ 'a hooked or U-shaped device used for gripping heavy objects' happen to be **homophones** or **homonyms**, words that sound alike but have unrelated meanings. However, a word doesn't have to be a homophone in order to be or belong to a lexeme. All words either are lexemes or belong to a lexeme's paradigm. So ROTATE, MOON, FAST, and GENTLY are all lexemes.

In order to talk about lexemes, morphologists give them each a name, and by convention, they put these names in capital or small capital letters. The name is generally the form by which a lexeme would be listed in a dictionary, and for this reason, we call it the **citation form**. The citation form is useful, but it does not necessarily have any mental status. Across languages, the citation form of a noun is most often the singular form. The citation form of a verb, however, varies widely. In English it is the bare infinitive (the infinitive minus "to": e.g., READ). In French and Spanish it is the infinitive (e.g., Fr. REGARDER, Sp. MIRAR 'to look at'). In Greek it is the first person singular form (e.g., MILO 'I speak').

In (11) each of the three lexemes had the same phonological shape but a different meaning. In (12) we see just the opposite: a single lexeme with more than one possible phonological shape. The verb *look* appears in several different forms, but with a consistent meaning:

(12) I *look*
she *looks*
we *looked*
they were *looking*

All of these forms are distinct phonological words, for the simple reason that they do not each sound exactly alike. In addition, we can label

them as different grammatical words (discussed in 2.1.3) because each plays a distinct grammatical role within a sentence. But at some level these different words are all tokens of the same type: they mean the same thing and no one would expect a dictionary to give them four separate entries. We must be dealing with a single lexeme, but one that happens to be realized in several different forms, depending on grammatical context. This example shows you that a lexeme is not a single form, but rather a set of forms. **[Exercises 6–7]**

What is a lexeme?

- A lexeme is a theoretical construct that corresponds roughly to one of the common senses of the term *word*. Examples include BOOK, EAT, DARK, SECRETLY.
- It is a sign or set of signs that exists independently of any particular syntactic context.
- It has a particular meaning or grammatical function (e.g., 'a set of written or printed pages fastened along one side and encased between two covers'; 'consume, as with food').
- Some linguists restrict the class of lexemes to the major lexical categories of noun, verb, adjective/adverb.
- It is generally referred to by its citation form (e.g., BOOK, EAT), but its shape may vary systematically according to the syntactic context in which it is used (e.g., one *book*, two *books*; I am *eating* right now, I *ate* a big dinner yesterday).

We have already said that in order to talk about lexemes, morphologists give them each a name, and by convention, they put these names in small capital letters. In reality the name of a lexeme is much more than a name. In English it also happens to be the lexeme's **lexical stem**. The lexical stem is the form of the lexeme that is most often used in the creation of new words.

To illustrate what we mean by lexical stem, let's look closely at the lexeme GO. This lexeme has five forms, two of them irregular: *go*, *goes*, *went*, *gone*, *going*. Of these forms, *go* has a different status from the rest. Lexemes formed from GO most often use it as their stem, as opposed to an inflected form. You have probably heard the word *churchgoer*, but not *church-wenter*, someone who used to go to church. Likewise, there are *go-betweens* but not *gone-betweens*.

It would be a mistake to overgeneralize and say that the lexical stem is always used in creating new words. We could use the word *went-between* and people would understand it. A quick glance at the dictionary reveals the forms *goner*, *going-over*, and *goings-on*, derived from the past and present participles *gone* and *going*. The last two examples are phrasal items – phrases that have been turned into words.

HAVE is the name of another lexeme that works the same way as GO. It has several distinct forms, some of them irregular: *have*, *has*, *had*, and *having*. Only one of them, HAVE, is generally used in forming new lexemes. The *haves and the have-nots* is a common expression, but not the *had-nots*, people who used not to have any money. (As an exercise, though, do an online search for *had-nots*. What sorts of examples do you find?) We generally do not make up words from the inflected forms *has*, *had*, or *having*. Exceptions, like *has-been*, are most likely to be phrasal items.

Nouns have two forms, a singular and a plural. The singular form is the lexical stem. It is therefore the form that most often appears in compounds. We say *apple-corer*, *boathouse*, *saber-tooth*, and *songwriter*, but never **apples-corers*, **boatshouse*, **saber-teeth*, or **songswriter*. This is only a generalization: occasionally we do find the plural form in a compound. Thus we have seen *antiques store*, *admissions office*, *customs house*, and *sports page*.

So that you won't get away thinking that all languages work like English in having one lexical stem per lexeme, consider the case of Latin. Latin has the peculiarity of having words, notably verbs, with more than one lexical stem. One lexical stem of the verb 'sing'[5] is *can-*. This is the stem found in *canō* 'I sing'. Another lexical stem is *cant-*, also called the participial stem. To form the word meaning 'singer', we take the participial stem *cant-* and the agentive suffix *-or*, giving the form *cantor*. The participial stem is always used in the formation of agentives.

To summarize, a lexeme is an abstract object, not a single concrete word, but a set of grammatical words. Cross-linguistically, one of those words is generally privileged to be the lexical stem from which other words are formed, although some languages permit more than one lexical stem. However, in morphology it is often safer to talk of tendencies than absolutes. Many phenomena are not categorical, but graded. So it is with the creation of new words from lexemes. Occasionally, particularly in the case of phrasal items like *has-been*, a form of a lexeme other than the lexical stem is used for creating new words. **[Exercise 8]**