Introducing Morphology

ROCHELLE LIEBER
English Department
University of New Hampshire

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521895491

First published in print format 2009

ISBN-13 978-0-511-77018-0 eBook (NetLibrary)

ISBN-13 978-0-521-89549-1 Hardback

ISBN-13 978-0-521-71979-7 Paperback

Contents

Preface

One of the things that drew me to linguistics several decades ago was a sense of wonder at both the superficial diversity and the underlying commonality of languages. My wonder arose in the process of working through my first few problem sets in linguistics, not surprisingly, problem sets that involved morphological analysis. What I learned first was not theory – indeed at that moment in linguistic history morphology was not perceived as a separate theoretical area in the US – but what languages were like, how to analyze data, and what to call things. I love morphological theory, but for drawing beginning students into the field of linguistics, I believe that there is no substitute for hands-on learning, and that is where this book starts.

This book is intended for undergraduate students who may have had no more than an introductory course in linguistics. It assumes that students know the International Phonetic Alphabet, and have a general idea of what linguistic rules are, but it presupposes little else in the way of sophistication or technical knowledge. It obviously assumes that students are English-speakers, and therefore the first few chapters concentrate on English, and to some extent on languages that are likely to be familiar to linguistics students from language study in high school and university. As the book progresses, I introduce data from many languages that will be "exotic" to students, so that by the end of the book, they will have some sense of linguistic diversity, at least with respect to types of morphology.

There are some aspects of the content of this text that might seem unusual to instructors. The first is the attention to dictionaries in chapter 2. Generally, texts on linguistic morphology do not mention dictionaries, but I find that beginning students of morphology retain a reverence for dictionaries that sometimes gets in the way of thinking about the nature of the mental lexicon and how word formation works. Instructors can skip all or part of this chapter, but my experience is that it sets students on a good footing from the start, and largely eliminates their squeamishness about considering whether *incent*, or *bovineness* or *organizationalize* or the like are 'real' words, even if we can't find them in the dictionary.

Another section that might seem odd is the part of chapter 7 devoted to snapshot descriptions of five different languages. These also might be skipped over, but they serve two important purposes. One purpose is simply to expose students to what the morphology of a language looks like overall; much of what they're exposed to in the rest of the book (and in most other morphology texts that I know of) are bits and pieces of the morphology of languages – a reduplication rule here, an inflectional paradigm there – but never the big picture. More importantly, having looked at the 'morphological toolkits' of several languages, students will be better prepared to understand both the traditional categories used in morphological typology and more recent means of classification.

The final thing that might strike instructors as unusual is that I largely hold off on introducing morphological theory until the last chapter. Clearly, no text is theory-neutral, and this text is no exception. It fits squarely in the tradition of generative morphology in the sense that I present morphology as an attempt to characterize and model the mental lexicon. I presuppose that there is much that is universal in spite of apparent diversity. And I believe that the ultimate aim of teaching students about morphology (indeed about any area of linguistics) is to expose them to what is at stake in trying to characterize the nature of the human language capacity. Nevertheless I start by presenting morphological rules in as neutral a way as possible, and hold off on raising theoretical disputes until students have enough experience to understand how

morphological data might support or refute theoretical hypotheses. In a sense I believe that students will gain a better understanding of theory if they already have the ability to find data and analyze it themselves. Therefore the bulk of the morphological theory will be found in the last chapter, where I have tried to pick a few theoretical debates and show how one might argue for or against particular analyses. Having read this chapter, students will be able to go on and tackle some of the texts that are intended for advanced undergraduates or graduate students.

Since one of my main goals in this text is to teach students to do morphology, there are a number of pedagogical features that set this book apart from other morphology texts. First, each chapter has one or more 'Challenge' boxes. These occur at points in the text where students might take a breather from reading or class lecture and try something out for themselves. Challenge exercises are ideal for small teams of students – either outside of class, or as an in-class activity – to work on together. Some involve discussion, some analysis, some doing some work on-line or at the library. But all of them involve hands-on learning. Instructors can use them or skip them or assign them as homework instead of, or in addition to, the exercises at the ends of chapters. I have tried most of them myself as in-class activities, and have found that they get students excited, stimulate discussion, and generally give students the feeling of really 'doing morphology' rather than just hearing about it.

A second pedagogical feature that sets this book apart are the "How to" sections in chapters 3, 4, 6, and 9. These are meant to give students tips on finding or working with data. Some students don't need such tips; they have the intuitive ability to look at data and figure out what to do with it. But I've found over years of teaching that there are some students who don't have this knack, and who benefit enormously from being walked through a problem or technique systematically. The "How to" sections do this.

Instructors and students will also find what they would expect to find in any good text. First, there are several aids to navigating the text – chapter outlines and lists of key terms at the beginnings of chapters and brief summaries at the end, as well as a glossary of the terms that are highlighted in the text. A copy of the International Phonetic Alphabet is included at the beginning for easy reference. And each chapter has a number of exercises that allow students to practice what they've been exposed to.

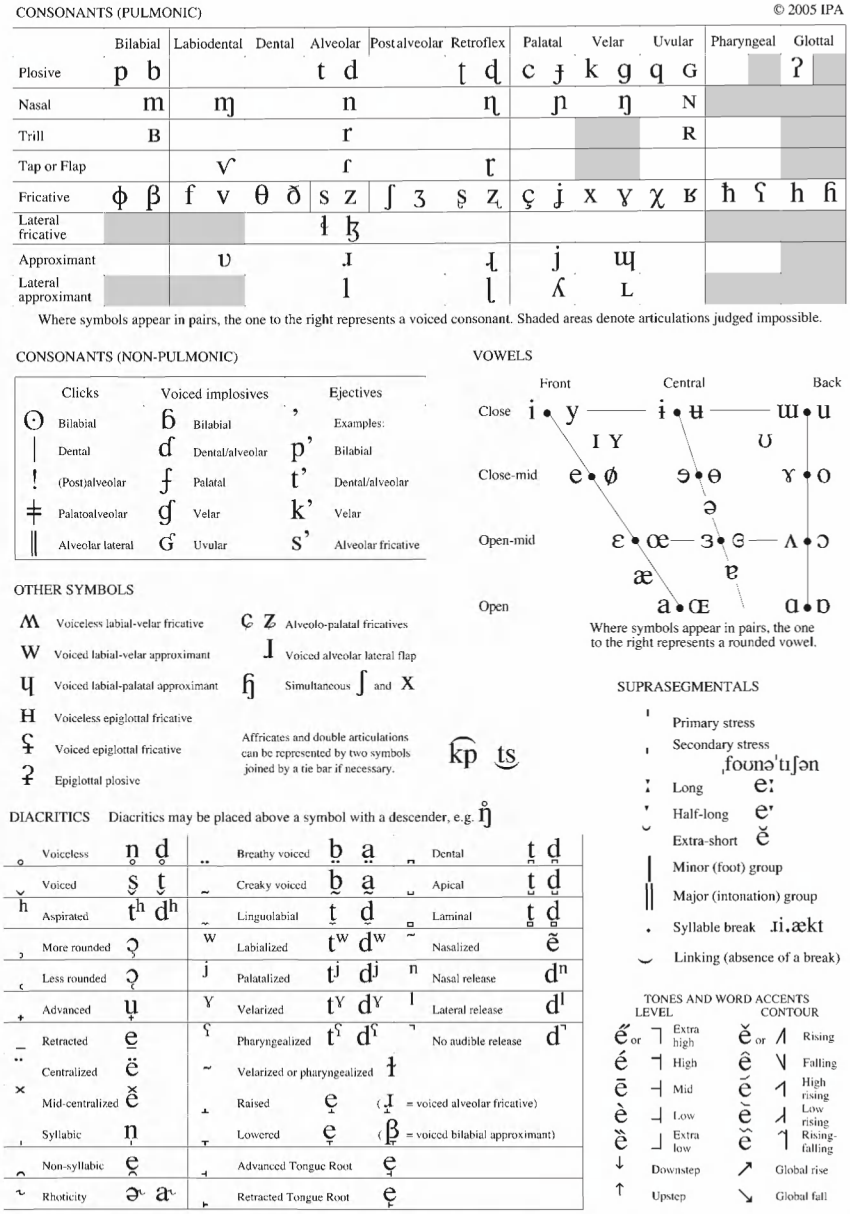
A general point about examples in this text. Where I have cited data from different books, grammars, dictionaries, and scholarly articles, I have chosen to keep the glosses provided in the original source even if this results in some inconsistency in the use of abbreviations. In other words, slightly different abbreviations may occur in different examples (for instance, N or Neut for 'neuter'). Although students may be confused by this practice at first, it does give them a taste of the linguistic "real world." Any student going on and doing further work in morphology is bound to find exactly this sort of variation in the use of abbreviations in sources.

My goal in this text is to bring students to the point where they are not only ready to confront morphological theory but also have the skills to begin to think independently about it, and perhaps to contribute to it.

This text has benefitted from the help of many people. I am grateful to John McCarthy and Donca Steriade for suggesting examples, to Charlotte Brewer for supplying me with statistics about citations in the OED, to Marianne Mithun for suggesting Nishnaabemwin as a polysynthetic language to profile, and to several classes of students at UNH both for serving as guinea pigs on early drafts and for supplying me with wonderful examples from their Word Logs. Thanks go as well to the College of Liberal Arts at the University of New Hampshire for the funds to hire a graduate student assistant at a critical moment, and to Chris Paris for supplying assistance. I am especially grateful to several anonymous reviewers who made excellent suggestions on the penultimate draft of the text. Finally, thanks are due as well to Andrew Winnard at Cambridge University Press for inviting me to write this text and for his patience in waiting for it.

The International Phonetic Alphabet

(revised to 2005)



Reproduced with the kind permission of the International Phonetic Association (Department of Theoretical and Applied Linguistics, School of English, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece).

Point and manner of articulation of English consonants and vowels

Consonants								
	Labial	Labio-dental	Interdental	Alveolar	Alveo-palatal	Palatal	Velar	Glottal
Stop	p,**b**			t,**d**			k,**g**	ʔ
Fricative		f,**v**	θ,ð	**s,z**	ʃ,ʒ			h
Affricate					ʧ,ʤ			
Nasal	**m**			**n**			ŋ	
Liquid				**ɹ,l**				
Glide	(**w**)					**j**	(**w**)	

Characters in boldface are voiced.
[w] is labio-velar in articulation.

Vowels			
	Front	Central	Back
High	i ɪ		u ʊ
Mid	e ɛ	ʌ,ə	o ɔ
Low	æ		ɑ

Tense vowels: i, e, u, o, ɑ
Lax vowels: ɪ, ɛ, æ, ʊ, ɔ, ʌ
Reduced vowel: ə

CHAPTER

1 What is morphology?

CHAPTER OUTLINE

KEY TERMS

morpheme
simplex
complex
type
token
lexeme
word form
inflection
derivation

In this chapter you will learn what morphology is, namely the study of word formation.

- We will look at the distinction between words and morphemes, between types, tokens, and lexemes and between inflection and derivation.
- We will also consider the reasons why languages have morphology.

1.1 Introduction

The short answer to the question with which we begin this text is that **morphology** is the study of word formation, including the ways new words are coined in the languages of the world, and the way forms of words are varied depending on how they're used in sentences. As a native speaker of your language you have intuitive knowledge of how to form new words, and every day you recognize and understand new words that you've never heard before.

Stop and think a minute:

- Suppose that *splinch* is a verb that means 'step on broken glass'; what is its past tense?
- Speakers of English use the suffixes *-ize* (*crystallize*) and *-ify* (*codify*) to form verbs from nouns. If you had to form a verb that means 'do something the way ex-Prime Minister Tony Blair does it', which suffix would you use? How about a verb meaning 'do something the way ex-President Bill Clinton does it'?
- It's possible to *rewash* or *reheat* something. Is it possible to *relove*, *reexplode*, or *rewiggle* something?

Chances are that you answered the first question with the past tense *splinched* (pronounced [splɪnʧt])[1], the second with the verbs *Blairify* and *Clintonize*, and that you're pretty sure that *relove*, *reexplode*, and *rewiggle* are weird, if not downright impossible. Your ability to make up these new words, and to make judgments about words that you think could never exist, suggests that you have intuitive knowledge of the principles of word formation in your language, even if you can't articulate what they are. Native speakers of other languages have similar knowledge of their languages. This book is about that knowledge, and about how we as linguists can find out what it is. Throughout this book, you will be looking into how you form and understand new words, and how speakers of other languages do the same. Many of our examples will come from English – since you're reading this book, I assume we have that language in common – but we'll also look beyond English to how words are formed in languages with which you might be familiar, and languages which you might never have encountered before. You'll learn not only the nuts and bolts of word formation – how things are put together in various languages and what to call those nuts and bolts – but also what this knowledge says about how the human mind is organized.

The beauty of studying morphology is that even as a beginning student you can look around you and bring new facts to bear on our study. At this point, you should start keeping track of interesting cases of new words

1. In this text I presuppose that you have already learned at least that part of the International Phonetic Alphabet (IPA) that is commonly used for transcribing English. You'll find an IPA chart at the beginning of this book, if you need to refresh your memory.

that you encounter in your life outside this class. Look at the first Challenge box.

Challenge: your word log

Keep track of every word you hear or see (or produce yourself) that you think you've never heard before. You might encounter words while listening to the radio, watching TV, or reading, or someone you're talking to might slip one in. Write those new words down, take note of where and when you heard/read/produced them, and jot down what you think they mean. What you write down may or may not be absolutely fresh new words – they just have to be new to you. We'll be coming back to these as the course progresses and putting them under the microscope.

Of course, if the answer to our initial question were as simple as the task in the box, you might expect this book to end right here. But there is of course much more to say about what makes up the study of morphology. Simple answers frequently lead to further questions, and here's one that we need to settle before we go on.

1.2 What's a word?

Ask anyone what a word is and ... they'll look puzzled. In some sense, we all know what words are – we can list words of various sorts at the drop of a hat. But ask us to define explicitly what a word is, and we're flummoxed. Someone might say that a word is a stretch of letters that occurs between blank spaces. But someone else is bound to point out that words don't have to be written for us to know that they're words. And in spoken (or signed) language, there are no spaces or pauses to delineate words. Yet we know what they are. Still another person might at this point try an answer like this: "A word is something small that means something," to which a devil's advocate might respond, "But what do you mean by 'something small'?" This is the point at which it becomes necessary to define a few specialized linguistic terms.

Linguists define a morpheme as the smallest unit of language that has its own meaning. Simple words like *giraffe*, *wiggle*, or *yellow* are morphemes, but so are prefixes like *re-* and *pre-* and suffixes like *-ize* and *-er*.[2] There's far more to be said about morphemes – as you'll see in later chapters of this book – but for now we can use the term **morpheme** to help us come up with a more precise and coherent definition of word. Let us now define a **word** as one or more morphemes that can stand alone in a language. Words that consist of only one morpheme, like the words in (1), can be

2. In chapter 2 we will give a more formal definition of **prefix** and **suffix**. For now it is enough to know that they are morphemes that cannot stand on their own, and that prefixes come before, and suffixes after, the root or main part of the word.

termed **simple** or **simplex** words. Words that are made up of more than one morpheme, like the ones in (2), are called **complex**:

(1) *Simplex words*
giraffe
fraud
murmur
oops
just
pistachio

(2) *Complex words*
opposition
intellectual
crystallize
prewash
repressive
blackboard

We now have a first pass at a definition of what a word is, but as we'll see, we can be far more precise.

1.3 Words and lexemes, types and tokens

How many words occur in the following sentence?

My friend and I walk to class together, because our classes are in the same building and we dislike walking alone.

You might have thought of at least two ways of answering this question, and maybe more. On the one hand, you might have counted every item individually, in which case your answer would have been 21. On the other hand, you might have thought about whether you should count the two instances of *and* in the sentence as a single word and not as separate words. You might even have thought about whether to count *walk* and *walking* or *class* and *classes* as different words: after all, if you were not a native speaker of English and you needed to look up what they meant in the dictionary, you'd just find one entry for each pair of words. So when you count words, you may count them in a number of ways.

Again, it's useful to have some special terms for how we count words. Let's say that if we are counting every instance in which a word occurs in a sentence, regardless of whether that word has occurred before or not, we are counting **word tokens**. If we count word tokens in the sentence above, we count 21. If, however, we are counting a word once, no matter how many times it occurs in a sentence, we are counting **word types**.

Counting this way, we count 20 types in the sentence above: the two tokens of the word *and* count as one type. A still different way of counting words would be to count what are called **lexemes**. **Lexemes** can be thought

of as families of words that differ only in their grammatical endings or grammatical forms; singular and plural forms of a noun (*class, classes*), present, past, and participle forms of verbs (*walk, walks, walked, walking*), different forms of a pronoun (*I, me, my, mine*) each represent a single lexeme. One way of thinking about lexemes is that they are the basis of dictionary entries; dictionaries typically have a single entry for each lexeme. So if we are counting lexemes in the sentence above, we would count *class* and *classes*, *walk* and *walking*, *I* and *my*, and *our* and *we* as single lexemes; the sentence then has 16 lexemes.

1.4 But is it *really* a word?

In some sense we now know what words are – or at least what word types, word tokens, and lexemes are. But there's another way we can ask the question "What's a word?" Consider the sort of question you might ask when playing Scrabble: "Is *aalii* a word?" Or when you encounter an unfamiliar word: "Is *bouncebackability* a word?" What you're asking when you answer questions like these, is really the question "Is *xyz* a REAL word?" Our first impulse in answering those questions is to run for our favorite dictionary; if it's a real word it ought to be in the dictionary.

But think about this answer for just a bit, and you'll begin to wonder if it makes sense. Who determines what goes in the dictionary in the first place? What if dictionaries differ in whether they list a particular word? For example, the *Official Scrabble Player's Dictionary* lists *aalii* but not *bouncebackability*. The *Oxford English Dictionary On-Line* doesn't list *aalii*, but it does list *bouncebackability*. So which one is right? Further, what about words like *cot potato* or *freshmore* that don't occur in any published dictionary yet, but can be encountered in the media? The former, according to Word Spy (www.wordspy.com) means a baby who spends too much time watching television (Americans might use the term *crib potato* instead of *cot potato*), and the latter is a second-year high school student in the US who has to repeat a lot of first-year classes. And what about the word *cot potatodom*, which I just made up? Once you know what a *cot potato* is, you have no trouble understanding my new word. If it consists of morphemes, has a meaning, and can stand alone, doesn't it qualify as a word according to our definition even if it doesn't appear in the dictionary?

What all these questions suggest is that we each have a **mental lexicon**, a sort of internalized dictionary that contains an enormous number of words that we can produce, or at least understand when we hear them. But we also have a set of **word formation rules** which allows us to create new words and understand new words when we encounter them. In the chapters to follow, we will explore the nature of our mental lexicon in detail, and think further about the "Is it really a word?" question. In answering this question we'll be led to a detailed exploration of the nature of our mental lexicon and our word formation rules.

1.5 Why do languages have morphology?

As native speakers of a language we use morphology for different reasons. We will go into both the functions of morphology and means of forming new words in great depth in the following chapters, but here, we'll just give you a taste of what's to come.

One reason for having morphology is to form new lexemes from old ones. We will refer to this as **lexeme formation**. (Many linguists use the term **word formation** in this specific sense, but this usage can be confusing, as all of morphology is sometimes referred to in a larger sense as 'word formation'.) Lexeme formation can do one of three things. It can change the part of speech (or **category**) of a word, for example, turning verbs into nouns or adjectives, or nouns into adjectives, as you can see in the examples in (3):

(3) *Category-changing lexeme formation*[3]
V→ N: amuse → amusement
V → A: impress → impressive
N → A: monster → monstrous

Some rules of lexeme formation do not change category, but they do add substantial new meaning:

(4) *Meaning-changing lexeme formation*

A → A 'negative A'	happy → unhappy
N → N 'place where N lives'	orphan → orphanage
V → V 'repeat action'	wash → rewash

And some rules of lexeme formation both change category and add substantial new meaning:

(5) *Both category and meaning-changing lexeme formation*

V → A 'able to be Ved'	wash → washable
N → V 'remove N from'	louse → delouse

Why have rules of lexeme formation? Imagine what it would be like to have to invent a wholly new word to express every single new concept. For example, if you wanted to talk about the process or result of amusing someone, you couldn't use *amusement*, but would have to have a term like *zorch* instead. And if you wanted to talk about the process or result of resenting someone, you couldn't use *resentment*, but would have to have something like *plitz* instead. And so on. As you can see, rules of lexeme formation allow for a measure of economy in our mental lexicons: we can recycle parts, as it were, to come up with new words. It is probably safe to say that all languages have some ways of forming new lexemes, although,

3. The notation V → N means 'changes a verb to a noun.'

as we'll see as this book progresses, those ways might be quite different from the means we use in English.

On the other hand, we sometimes use morphology even when we don't need new lexemes. For example, we saw that each lexeme can have a number of **word forms**. The lexeme WALK has forms like *walk, walks, walked, walking* that can be used in different grammatical contexts. When we change the form of a word so that it fits in a particular grammatical context, we are concerned with what linguists call **inflection**. Inflectional word formation is word formation that expresses grammatical distinctions like number (singular *vs.* plural); tense (present *vs.* past); person (first, second, or third); and case (subject, object, possessive), among others. It does not result in the creation of new lexemes, but merely changes the grammatical form of lexemes to fit into different grammatical contexts.

Interestingly, languages have wildly differing amounts of inflection. English has relatively little inflection. We create different forms of nouns according to number (*wombat, wombats*); we mark the possessive form of a noun with *-'s* or *-s'* (*the wombat's eyes*). We have different forms of verbs for present and past and for present and past participles (*sing, sang, singing, sung*), and we use a suffix *-s* to mark the third person singular of a verb (*she sings*).

However, if you've studied Latin, Russian, ancient Greek, or even Old English, you'll know that these languages have quite a bit more inflectional morphology than English does. Even languages like French and Spanish have more inflectional forms of verbs than English does.

But some languages have much less inflection than English does. Mandarin Chinese, for example, has almost none. Rather than marking plurals by suffixes as English does, or by prefixes as the Bantu language Swahili does, Chinese does not mark plurals or past tenses with morphology at all. This is not to say that a speaker of Mandarin cannot express whether it is one giraffe, two giraffes, or many giraffes that are under discussion, or whether the sighting was yesterday or today. It simply means that to do so, a speaker of Mandarin must use a separate word like *one, two* or *many* or a separate word for *past* to make the distinction.

(6) Wo jian guo yi zhi chang jing lu.
I see past one CLASSIFIER giraffe[4]

(7) Wo jian guo liang zhi chang jing lu
I see past two CLASSIFIER giraffe

The word *chang jing lu* 'giraffe' has the same form regardless of how many long-necked beasts are of interest. And the verb 'to see' does not change its form for the past tense; instead, the separate word *guo* is added to express this concept. In other words, some concepts that are expressed via inflection in some languages are expressed by other means (word order, separate words) in other languages.

4. We will explain in chapter 6 what we mean by **classifier**. For now it is enough to know that classifiers are words that must be used together with numbers in Mandarin.

1.6 The organization of this book

In what follows, we'll return to all the questions we've raised here. In chapter 2, we'll revisit the question of what a word is, by further probing the differences between our mental lexicon and the dictionary, and look further into questions of what constitutes a "real" word. We'll look at the ways in which word formation goes on around us all the time, and consider how children (and adults) acquire words, and how our mental lexicons are organized so that we can access the words we know and make up new ones. In chapter 3, we'll get down to the work of looking at some of the most common ways that new lexemes are formed: by adding prefixes and suffixes, by making up compound words, and by changing the category of words without changing the words themselves. In this chapter we'll concentrate on how words are structured in terms of both their forms and their meanings. Many of our examples will be taken from English, but we'll also look at how these kinds of word formation work in other languages. Chapter 4 takes up a related topic, productivity: some processes of word formation allow us to form many new words freely, but others are more restricted. In this chapter we'll look at some of the determinants of productivity, and how productivity can be measured. Chapter 5 will also be concerned with lexeme formation, but with kinds of lexeme formation that are less familiar to speakers of English. We'll look at forms of affixation that English does not have (infixation, circumfixation), processes like reduplication, and templatic morphology. Our focus will be on learning to analyze data that might on the surface seem to be quite unfamiliar. In chapter 6 we will turn to inflection, looking not only at the sorts of inflection we find in English and other familiar languages, but also at inflectional systems based on different grammatical distinctions than we find in English, and systems that are far more complex and intricate. Chapter 7 will be devoted to the subject of typology, different ways in which the morphological systems of the languages of the world can be classified and compared to one another. We'll look at some traditional systems of classification, as well as some that have been proposed more recently, and assess their pros and cons. Chapters 8 and 9 will explore the relationship between the field of morphology and the fields of syntax on the one hand and phonology on the other. Our final chapter will introduce you to some of the interesting theoretical debates that have arisen in the field of morphology over the last two decades and prepare you to do more advanced work in morphology.

Summary

Morphology is the study of words and word formation. In this chapter we have considered what a word is and looked at the distinction between word tokens, word types, and lexemes. We have divided word formation into derivation – the formation of new lexemes – and inflection, the different grammatical word forms that make up lexemes.

Exercises

1. Are the following words simple or complex?
   a. members
   b. prioritize
   c. handsome
   d. fizzy
   e. dizzy
   f. grammar
   g. writer
   h. rewind
   i. reject
   j. alligator

   If you have difficulty deciding whether particular words are simple or complex, explain why you find them problematic.
2. Do the words in the following pairs belong to the same lexeme or to different lexemes?
   a. revolve — revolution
   b. revolution — revolutions
   c. revolve — dissolve
   d. go — went
   e. wash — rewash
3. In the following sentences, count word tokens, types, and lexemes:
   a. I say now, just as I said yesterday, that the price of a wombat is high but the price of a platypus is higher.

      tokens _____
      types _____
      lexemes _____

   b. I've just replaced my printer with a new one that prints much faster.

      tokens _____
      types _____
      lexemes _____

4. In sentence (3b), what sorts of problems does the word *I've* pose for our definition of 'word'?
5. What words belong to the same word family or lexeme as *sing*?

CHAPTER

2 Words, dictionaries, and the mental lexicon

CHAPTER OUTLINE

KEY TERMS

word
mental lexicon
lexicography
the Gavagai problem
fast mapping
aphasia

In this chapter you will learn why we make a basic distinction between the dictionary and the mental lexicon.

- We will look at how linguists study the mental lexicon and how children acquire words.
- We will consider whether complex words are stored in the mental lexicon, or derived by rules, or both.
- And we will look further at how dictionaries have evolved and how they differ from one another and from the mental lexicon.

2.1 Introduction

In the last chapter, we raised the question "what's a word?" And we saw in section 1.2 that this question actually subsumes two more specific questions. In this chapter we will look more closely at those questions.

On the one hand, when we ask "what's a word?," we may be asking about the fundamental nature of wordhood – as we saw, a far thornier philosophical question than it would seem at first blush. Native speakers of a language seem to know intuitively what a 'word' is in their language, even if they have trouble coming up with a definition of 'word'. Interestingly, the *Oxford American Dictionary* seems to bank on this intuitive knowledge when it defines a word as "a single distinct meaningful element of speech or writing, used with others (or sometimes alone) to form a sentence and typically shown with a space on either side when written or printed." We've already debunked part of the OAD definition: languages need not be written, but they still have words, and words don't have blank space between them in spoken language. Nevertheless, the OAD's definition works for most people: most dictionary users probably do not know the word **morpheme**, which we used in our definition of word in the last chapter, but the OAD relies on the likelihood that they will not first think of something like the prefix *re-* as a single meaningful element, or something like *irniarualiunga* which means 'I am making a doll' in Central Alaskan Yup'ik (Mithun 1999: 203), and constitutes not only a word, but also a whole sentence. In other words, the OAD's definition works because dictionary users already have an intuitive idea of what a word is!

Morphologists, however, have the luxury of being more precise: we can define a word as a sequence of one or more morphemes that can stand alone in a language. But in doing so, we have not exhausted what's interesting about our question.

Indeed, in chapter 1 we saw that there is a second way of interpreting it, one that seems far more concrete at first: we can interpret our question as meaning "Is *xyz* a word?" where *xyz* is a specific morpheme or sequence of morphemes. Taken this way, our question asks what it means to say that *xyz* is a word of English, or Central Alaskan Yup'ik, or some other language. On the one hand, we are always making up new words, and when we say them, others understand what we mean. In the last chapter, I mentioned the words *freshmore* and *cot potatodom*, neither of which is in a (conventional) dictionary, at least as of the writing of this chapter, but both of which have been used (at least by me!). Does this qualify them as words? And two paragraphs up, I used the word *wordhood*, which you may or may not like, but which you certainly understood. This is the version of the "what's a word" question that we'll concentrate on in this chapter. In doing so we'll begin to explore the nature of dictionaries, and more importantly of our native speaker knowledge of words, which we might term our mental lexicon.

2.2 Why not check the dictionary?

When the question "Is *xyz* really a word?" comes up – whether in casual conversation, in reading an article in the newspaper, or in playing Scrabble – people will often look to the dictionary for an answer. Which dictionary, of course, depends on what's lying around the house or the office, or these days, what's available on-line. But is this the right way to answer our question? As morphologists, we need to think about how dictionaries come to be, and how much we credit them with the authority to decide what's a word.

There's a lot to be said about how dictionaries have evolved and how they are produced today. For a short history of English dictionaries, you can read section 2.4 of this chapter. But for our immediate purposes, we can identify a number of reasons why we wouldn't always want to base the answer to our question on what we find (or don't) in a dictionary. Here are a few such reasons.

2.2.1 Which dictionary?

Dictionaries come in all shapes and sizes, for all sorts of intended audiences. Size and audience are determined by individual publishers, and indeed the finished product is shaped by all sorts of market forces. And makers of dictionaries – lexicographers – are of course human; what gets into dictionaries has historically been subject to the individual foibles of lexicographers, not to mention the mores of society. If you grew up when I did, it was typical for dictionaries not to have taboo words like *fuck*, much less its derivatives *fucking, fuck up, fuckable, fuck all,* and *fucker*, all of which can be found today in the *Concise Oxford English Dictonary*; but until the 1970s, dictionaries avoided words that might offend. It is perhaps safe to say that individual or societal foibles play less of a role in dictionary-making today, but it's still a good idea to keep in mind that neither lexicographers nor the dictionaries they create are infallible.

Our first problem with giving final authority for wordhood to the dictionary, then, follows from the very concrete and temporal nature of dictionaries: if you look up a word in a pocket dictionary, or even a standard college desk dictionary, and it isn't listed, you might still have the nagging suspicion that a bigger dictionary or a more specialized dictionary might list the word. But even if you check the largest available dictionary – say, for English the *Oxford English Dictionary On-line* – or the most complete technical dictionary in a particular field, can you be sure that a word that's not listed isn't a word? Maybe it's too new a word to have gotten into the dictionary yet.

2.2.2 Nonces, mistakes, and mountweazels

Further, sometimes we find items in dictionaries that we might hesitate to call words – even if they do occur in the dictionary. Among these items are words that are labeled as 'nonce', meaning that they've been found just once, often in the writing of someone important, but that nevertheless

don't seem to occur anywhere else. The OED On-line, for example, lists as a nonce the word *agreemony*, which they define as 'agreeableness', and illustrate with a single quotation from the seventeenth-century writer Aphra Behn. Was this ever really a word? Indeed, the OED even lists some words that occur only once, and further, in contexts which don't illuminate their meaning; for example, we can find the word *umbershoot* used by James Joyce in *Ulysses*, about which the OED maddeningly says only "meaning obscure"! Words or not?

Very extensive dictionaries like the OED sometimes also contain words that they identify as mistakes. For example, we can find an entry for the word *ambassady*, which occurs in a single quotation from 1693 and is, according to the OED, perhaps a mistake, where the author might have meant the word *ambassade* "the mission or function of an ambassador." It occurs in the dictionary, but is it really a word?

And finally, there are what have come to be called 'mountweazels'. A mountweazel is a phony word that is inserted into a dictionary so that its makers can identify lexicographic piracy. You can find a fuller explanation of this tradition in section 2.4, but the short version is this: lexicographers sometimes make up an entry and include it so that they can tell if another lexicographer is using their dictionary as a source without attribution (which is plagiarism, of course). Surely we wouldn't want to count such impostors as real words, but they're in the dictionary!

2.2.3 And the problem of complex words

We will learn much more about this in the chapters to come, but perhaps the worst problem for us with the idea of giving the dictionary the authority to determine whether *xyz* is a word is that dictionaries don't need to include every word. Every language has ways of forming new words that are so active and transparent that putting all the words formed that way into the dictionary would be a waste of space. For example, speakers of English know that any verb at all can have a present progressive form made with the suffix *-ing*. As soon as I make up a new verb, say *zax*, we know that the present progressive verb form is *zaxing*. So although a dictionary might eventually have to include the verb *zax*, it might never list *zaxing* as a word. But of course *zaxing* should be considered a word. Similarly, just about any adjective in English can be made into a noun by adding the suffix *-ness*. For example, the *Concise Oxford English Dictionary* contains the adjective *bovine*, but not the noun *bovineness*. Nevertheless, I'd have no problem if I saw the word *bovineness* written somewhere, and would never think to look it up in the dictionary. The dictionary doesn't have the word precisely because we'd never need to look it up.

The conclusion that we are inexorably led to is that we cannot rely on dictionaries to answer the question "Is *xyz* a word?" On the one hand, dictionaries don't list all the words of any language. They can't list all derivatives with living prefixes and suffixes, or all technical, scientific, regional, or slang words. And on the other hand, they sometimes include

words used only once whose meanings are completely unknown. They occasionally even include purposely made-up words to guard their own copyrights. For the most part, dictionaries do not fix or codify the words of a language, but rather reflect the words that native speakers use. Those words are encoded in what we will call the mental lexicon, the sum total of word knowledge that native speakers carry around in their heads. So to answer our question, we must look more closely at what is in that mental lexicon.

2.3 The mental lexicon

By the mental lexicon I mean the sum total of everything an individual speaker knows about the words of her language. This knowledge includes information about pronunciation, category (part of speech), and meaning, of course, but also information about syntactic properties (for example, whether a verb is transitive or intransitive), level of formality, and what lexicographers call 'range of application', that is, the specific conditions under which we might use the word. For example, I know that the word *verandah* is a noun, pronounced (in my American English) [vəɹ**æn**də],[1] that it refers to a type of porch, and that I'd only use it in reference to the sort of porch one finds in the southern part of the US or perhaps in some exotic tropical country. Unless I was being ironic, I probably would not call my own back porch 'the verandah'. I also know that *barf* is a verb that's pronounced [bɑɹf], that it means 'vomit', that it is intransitive (unless used with a particle like *up*) and that it is used only colloquially (I wouldn't use it if I were describing the symptoms of a stomach flu to the doctor).

It is quite likely that in our mental lexicons we have entries that are only partial. We may know the pronunciation of a word, but not its meaning (e.g., I know how to pronounce *amortize*, but I'm not sure what it means). Or the opposite: for example, I know what the word *hegemony* means, but I don't know if it's pronounced with the stress on the first or second syllable. We may also have only partial knowledge of the meaning of a word. I know, for example, that a *distributor* is part of a car and that if you have to replace it, it's a relatively expensive job, but I don't know what a distributor looks like or what it does.

Each person's mental lexicon is sure to contain things that are different from other people's mental lexicons. One person may know lots of words for types of birds or flowers, another might know all the specialized vocabulary of sailing, and so on. Auto mechanics surely know more details of the meaning of the word *distributor* than I do. But our individual mental lexicons overlap enough that we speak the same language. In this section we will look in more detail at the contents of our mental lexicons, both what is stored and what is created by rules of word formation, and how our mental lexicons are organized.

1. Stressed syllables are marked by bold type.

2.3.1 How many words?

Psycholinguists estimate that the average English-speaking six-year-old knows 10,000 words, and the average high-school graduate around 60,000 words. Paul Bloom describes how this estimate can be made (2000: 5):

> Words are taken from a large unabridged dictionary, including only those words whose meanings cannot be guessed using principles of morphology or analogy. . . . Since it would take too long to test people on hundreds of thousands of words, a random sample is taken. The proportion of the sample that people know is used to generate an estimate of their overall vocabulary size, under the assumption that the size of the dictionary is a reasonable estimate of the size of the language as a whole. For example, if you use a dictionary with 500,000 words, and test people on a 500-word sample, you would determine the number of English words they know by taking the number that they got correct from this sample and multiplying by 1,000.

Children generally begin to produce their first words around the age of one. Bloom calculates that between the ages of one and 18 we would have to learn approximately ten words every day to have a vocabulary of 60,000 words. It's worth pointing out, I think, that this figure just takes into account the words that we have stored (fully or partially) in our mental lexicon, and not the words – perhaps an infinite number of them – that we can create by using rules of word formation. We will return shortly to our knowledge of word formation rules and its relation to our mental lexicon. First, however, we will look more closely at how we acquire our mental lexicon.

2.3.2 The acquisition of lexical knowledge

Psycholinguists have devised experiments to try to learn how children and adults are able to acquire words so easily. You might think that the learning of new words is a simple matter of association: someone points at something and says "flurge" and you learn that that something is called a *flurge*. This may be the way that we learn some words, but surely not the way we learn the majority of words in our mental lexicons. For one thing, not everything for which we have a word can be pointed at.

And even if someone points and says a word, it is often not clear from the context what exactly is being pointed out. Psycholinguists sometimes call this **the Gavagai problem**, following a scenario first discussed by the philosopher W.O. Quine. To summarize:

> Picture yourself on a safari with a guide who does not speak English. All of a sudden, a large brown rabbit runs across a field some distance from you. The guide points and says "gavagai!" What does he mean?
>
> One possibility is, of course, that he's giving you his word for 'rabbit'. But why couldn't he be saying something like "There goes a rabbit running across the field"? or perhaps "a brown one," or "Watch out!," or even "Those are really tasty!"? How do you know?
>
> In other words, there may be so much going on in our immediate environment that an act of pointing while saying a word, phrase, or sentence will not determine clearly what the speaker intends his utterance to refer to.

Besides, we are rarely in a situation in which someone is actively instructing us about the meanings of words; although parents may point to things in a picture book and name them for a child, or school children may be asked to memorize a list of vocabulary words, we learn most words without explicit instruction and seemingly with very little exposure. Although we do not know nearly enough about this subject, there are several things that we do know about how word learning occurs.

First, it is believed that both children and adults are able to do what the psycholinguist Susan Carey has called fast mapping (Carey 1978). Fast mapping is the ability to pick up new words on the basis of a few random exposures to them. In one experiment, Carey showed that children who were casually exposed to a new color name *chromium* during an unrelated activity (following instructions to pick up trays of various colors) were able to absorb the word and recall it even six weeks later. Experiments have shown that adults exhibit this fast mapping ability as well; while the ability to learn linguistic rules (say, of syntax or phonology) is thought to decline after puberty, the ability to learn new words remains robust.

Challenge

Here's an experiment you can try. Collect five or six objects. All but one of your objects should be familiar items (a bunch of keys, a mug, a pencil, etc.). One object, however, should be something odd and not familiar to many people. Put all your objects on a tray, and ask your subject (anyone outside your class will do) to point out the *zorch*. Observe what you subject does. Now take away the unfamiliar object, leaving only the familiar objects, and ask a different subject to point out the *plitz*. Again, observe closely what the subject does.

Psycholinguists have proposed a number of other strategies that both children and adults seem to use in learning new words.[2] One might be called the **Lexical Contrast Principle**. For example, in an experiment similar to yours, children were asked to point to the *zorch* (or some other made-up word), and what they invariably did was to point out the unfamiliar object. According to the Lexical Contrast Principle, the language learner will always assume that a new word refers to something that does not already have a name.

A second word learning strategy might be called the **Whole Object Principle**. In the experimental condition described above, when subjects are presented with the word *zorch* and an unnamed object, they will assume the whole unnamed object to be a *zorch*. They will not assume that *zorch* refers to a part of the object, to its color or shape, or to a superordinate category of objects to which it might belong.

2. See Bloom (2000) for an extensive discussion of this subject.

A related strategy might be dubbed the **Mutual Exclusivity Principle**. In the second experiment above, there are only familiar objects for which subjects already have names. When asked to point out the *plitz*, experimental subjects typically do one of two things: they might first look around the room for something else that might be called a *plitz*, or they might assume that the word *plitz* refers to a part of one of the familiar objects or a special type of one of them. Subjects, in other words, will assume that if an object already has a word for it, the word *plitz* cannot be synonymous with those words.

These experiments are of course not just hypothetical. Paul Bloom, Susan Carey, and many other psycholinguists have conducted them both with children of various ages and with adults, and have obtained the results described above. What is perhaps most astonishing about their results is that their experimental subjects often remember the words they've been exposed to when they are retested weeks after the original experiment. But maybe we should not be surprised by this: how otherwise could we have learned 60,000 words by the time we're 18?

Children not only learn individual words, but – as we'll see in the chapters to come – they learn the rules that allow us to create and understand new words. Indeed, there is evidence that English-speaking children as young as 18- to 24-months old are able to create new compound words (that is, words like *wind mill* or *dog bed*) and to turn nouns into verbs, a process which is called conversion (see chapter 3). Not too long after this, children will begin to use prefixes and suffixes, both for inflection and lexeme formation. We know that they have learned the rules when they produce words that are novel and therefore that they could not have learned from the language spoken around them.

2.3.3 The organization of the mental lexicon: storage *versus* rules

Although linguists like to describe our knowledge of words as a mental lexicon, we know that the mental lexicon is not organized alphabetically like a dictionary. Rather, it is a complex web composed of stored items (morphemes, words, idiomatic phrases) that may be related to each other by the sounds that form them and by their meanings. Along with these stored items we also have rules that allow us to combine morphemes in different ways. Our evidence for this organization comes from experiments using both normal subjects and subjects with some sort of genetic disorder or trauma to the brain.

There is a great deal of evidence to support the idea that speakers do not merely learn and store complex words (although they may store some complex words which are used frequently), but rather construct complex words using rules of word formation. We will go into great detail in the chapters to come on exactly what these rules of word formation look like, but let us start with a simple example, and use that example to explore what linguist Steven Pinker calls the "words and rules" theory of the mental lexicon (Pinker 1999).

We will take as our example the rule for forming past tenses of verbs in English. At this point, if I asked you how to form the past tense of a verb

in English, you would probably say that you usually add an *-ed*. And then you might point out that there are a number of verbs that have irregular past tenses like *sing~sang, tell~told, win~won, fly~flew,* and the like. We will look first at the regular past tense rule.

While it is true that in writing we add an *-ed* to form the past tense of a verb, in terms of spoken speech, the situation is a bit more complicated. Consider the next Challenge:

Challenge

Consider how you *pronounce* the past tenses of these verbs:

1. rap, tack, laugh, sheath, pass, lurch
2. pat, prod
3. rob, rove, bathe, buzz, rouge, judge, warm, warn, bang, roar, rule, tango

Transcribe the past tenses of these words in the International Phonetic Alphabet and observe how they differ.

You pronounce the past tenses of the first set of words in the Challenge box with a [t] sound, in the second with a sound like [əd], and the third with a [d] sound.

We do not choose the pronunciation of the past tense at random. Rather, the choice of which of the three endings to use depends on the final sound of the verb. Those words that are pronounced with final [t] or [d] sounds – those in the second list – get the [əd] pronunciation. The words that end in voiceless (with the exception of [t]) sounds get the [t] pronunciation. And all the rest get the [d] pronunciation. As for irregular forms like *sang* and *flew*, we must assume that English speakers simply learn them as exceptions.

We know that speakers of English have an unconscious knowledge of the past tense rule because we can automatically create the past tense of novel verbs. For example, if I coin a verb *blick*, you know that the past tense morpheme is pronounced [t]. Similarly, the novel verb *flurd* will have the past tense [əd], and the verb *zove* will be made past tense with [d]. We can even form the past tense of verbs that contain final sounds that do not occur at all in English, and when we do, we still follow the rule. For example, if we imagine that there are many composers imitating the style of Johann Sebastian Bach, and we coin the verb *to bach* to denote the action of imitating Bach, we will automatically form the past tense with the past tense variant pronounced [t], because the final sound of Bach is [x], a voiceless velar fricative. The important point here is that when we hear this sound at the end of a verb we know (unconsciously) that it's voiceless, and apply the past tense rule to it in the usual way.

Now that we know something about the English past tense rule, we can return to the question of how the mental lexicon is organized. It might be plausible to assume that speakers of English use the past

tense rule when they are creating the past tenses of novel verbs, but simply store the past tense forms of words they have already heard. In other words, we might assume that once a past tense has been formed, it is entered whole in our mental lexicon, and we retrieve it whole just as we would the present tense form. This hypothesis, however, may not be correct.

2.3.4 Evidence from aphasia

Studies of aphasics – people whose language faculty has been impaired due to stroke or other brain trauma – show that there must be a past tense rule that speakers use for regular forms – even very frequent ones – and that irregular forms are stored whole, probably in a different part of the brain. Badecker and Caramazza (1999) describe how we can know this.

Some aphasics display **agrammatism**; this means that they have difficulty in producing or processing function words in sentences, but can still produce and understand content words. Interestingly, agrammatic aphasics have difficulty producing or processing both regularly inflected forms (like the English past tenses), and also productively derived words (those with suffixes that we use frequently in making up new words – for example, *-less* as in *shoeless* or *-ly* as in *darkly*), whereas they have far less trouble with irregular forms like *sang* and *flew*.

Other aphasics display **jargon aphasia**; these aphasics produce fluent sentences using function words, but have trouble producing and understanding content words. Instead, they have a tendency to produce nonsense words. Interestingly, jargon aphasics will use regular inflections appropriately on their nonsense words, but they have difficulty processing and producing irregular forms.

We can explain the differential behavior of agrammatical and jargon aphasics if we postulate that we have rules for producing regularly inflected and productively derived forms, and only store irregular forms, and that rules and stored items are located in different parts of the brain. For agrammatic aphasics, the rule is unavailable, presumably because the part of the brain has been damaged that apparently allows us to apply morphological rules, but the irregular forms are still accessible from an undamaged part of the brain. For jargon aphasics, the irregular forms have been lost because the part of the brain that apparently allows access to stored forms has been damaged, but the regular rule is still intact.

2.3.5 Evidence from imaging studies

Imaging studies of normal subjects, such as those done with **PET (positron emission tomography) scans** seem to show the same thing. PET scans measure the level of blood flow to different parts of the brain, which in turn shows us areas of activation in those parts. Jaeger *et al.* (1996) have reported that there are parts of the brain that are activated when subjects are asked to read regularly inflected past tenses that are distinct from those activated in reading or producing irregular past tenses.

2.3.6 Evidence from genetic disorders

Similar conclusions follow from studies of two different genetic disorders – **Specific Language Impairment (SLI)** and **Williams Syndrome** – that affect language in different ways. Individuals with SLI are generally of normal intelligence and have no hearing impairment. But they are slow to produce and understand language, and their speech is characterized by the omission of various inflectional morphemes. Individuals with Williams Syndrome have a genetic disorder linked to various heart problems, elevated levels of calcium in their blood, and a characteristic appearance (short stature, an upturned nose, a long neck, among other things). Their language and social skills are in the normal range, but in other respects such as motor control and spatial perception they display mild or moderate developmental delay.

What is significant for our purposes is that these disorders provide more evidence for the organization of our mental lexicon. Individuals with SLI find it difficult to create the past tenses of novel verbs, and often fail to inflect unfamiliar regular verbs correctly; they have less difficulty with irregular verbs, though. In spontaneous speech, they may leave the regular past tense off verbs (Redmond and Rice 2001). In contrast, individuals with Williams Syndrome speak fluently and produce sentences with correct regular past tenses, but have more trouble with irregular ones; indeed they seem to use regular past tense marking even where control subjects or individuals with SLI would not, for example, overgeneralizing the regular *-ed* ending on irregular verbs (for example, *falled*) (Clahsen, Ring, and Temple 2004). Assuming that the genetic anomalies associated with these disorders affect different parts of the brain, we can explain this pattern of behavior.

2.3.7 Reprise: is it really a word?

We have spent some time in this chapter contrasting the dictionary with the mental lexicon in order to understand the question "Is *xyz* really a word?" We are now in a position to understand this question better, at least from the point of view of morphologists. Most morphologists would say that *xyz* is a word if it can be formed by the rules of word formation in a particular language. So words like *wordhood* or *re-reprise* that you might never have seen before you read this chapter really are words, even though you won't find them in any dictionary. They are words because they follow the rules of English word formation. It is the rules of word formation that we know that most distinguish our mental lexicon from the dictionary. The dictionary does not need to list all the words that we know or that we could create, because once we know word formation rules we can produce and understand potentially infinite numbers of new words from the morphemes available to us. The remainder of this book will be an attempt to work out in some detail what those rules are.

2.4 More about dictionaries

In section 2.2 we considered all the reasons why morphologists don't look upon dictionaries as the ultimate arbiters of 'wordhood' in English, or indeed in any language. You may not need more convincing of this issue,

but for those of you who have a fondness for dictionaries (most morphologists do!), it's worth knowing something about how dictionaries have developed. I'll again concentrate on English here, as our common language, but the history of dictionary-making for other languages can be equally fascinating.

Let's start with a thought experiment. Look at the next Challenge.

Challenge

Suppose that a great catastrophe has occurred and every single written or on-line dictionary has disappeared from the face of the earth. You and your classmates have survived the catastrophe (perhaps in a hidden concrete tunnel beneath the building in which you are now sitting), and have been delegated the task by other survivors of creating the first post-catastrophe dictionary of your language.

How would you start?

Your first instinct would probably be to make a list of words that you would need to define. Assuming that there were no surviving books to use as dictionary-fodder, a good way to begin would be by thinking of categories, and listing everything you could in each one. After you've listed all the animals, plants, and types of furniture you could think of, you'd come up with a list of hairstyles (*crewcut, bob, beehive, bun, buzz cut, duck's ass, cornrows, mullet, . . .*) and condiments (*ketchup, soy sauce, mustard, horseradish, wasabi, sambal oelek, . . .*), and so on, and eventually you'd come to articles (*a, the, this, that, . . .*), prepositions (*in, on, above, during, for, . . .*) and the other small words that form the grammatical glue that holds sentences together.

But along the way, you'd discover a number of problems. First, you'd have a suspicion that you'd be forgetting things (what, for example, was the name for that women's hairstyle that was the rage in the seventies?). Second, you and your classmates would get into constant arguments over this word or that: is it worth putting the word *mullet* in the dictionary as the name of a hairstyle? Wasn't that slang? Does slang go in the dictionary? What IS slang, anyway? Is it too vulgar to put *duck's ass* in the dictionary as a name of a 1950s hairstyle? What about really raunchy words? Is *sambal oelek* a word for a condiment in English, or is it just something we've borrowed from another language (what other language, though?)?

What this thought experiment does is to put you in the shoes of a lexicographer. In reality, it's been centuries since lexicographers have had to start from scratch in creating a dictionary – and perhaps they've never really done so. As the lexicographer Sidney Landau has said about the tradition of dictionary-making in English (2001: 43), "The history of English lexicography usually consists of a recital of successive and often successful acts of piracy." For years and years, each succeeding dictionary-maker has consulted already existing dictionaries to come up with a base

list of words, often adding new ones and sometimes deleting words for various reasons. But at least at first, lexicographers did have to decide one by one on each of the English words to include. Of course, there were manuscripts and books available to suggest words that needed to be included, and in fact, the earliest English lexicographers did rely on the words they found in books as the material from which they built their dictionaries.

2.4.1 Early dictionaries

It was not until the early seventeenth century that anything we would recognize as a monolingual dictionary could be found for the English language. Dictionaries or glossaries for translating Latin to English date from a century or so earlier, and in the sixteenth century lists of so-called hard words could be found for English, explaining words which largely had been adapted from Latin. The first real dictionary of English is generally acknowledged to be Robert Cawdrey's (1604) *A Table Alphabeticall of Hard Words*. The tradition of lexicographical piracy goes at least as far back as Cawdrey, who is said to have used an available Latin–English dictionary of his day to help come up with the words to define. The first dictionaries going beyond the tradition of defining only 'hard' words to include ordinary, everyday words began to appear in the early eighteenth century; Landau (2001: 52) cites John Kersey's (1702) *A New English Dictionary* as the earliest of these, followed by Nathaniel Bailey's (1721) *An Universal Etymological English Dictionary*.

2.4.2 Johnson's dictionary

A more significant milestone in the history of English lexicography for our purposes was Samuel Johnson's *Dictionary of the English Language*, published in 1755. It contains more than 42,000 entries – even then, only a small fraction of English vocabulary – and took seven years to write, an astonishing feat for a single individual. Johnson's dictionary was not only the most comprehensive English dictionary of his time, but it was also among the first dictionaries to include illustrative quotations on a large scale.

What is most interesting for our purposes, though, are the idiosyncrasies of Johnson's dictionary: what he included, what he left out, and how he defined various words in odd ways. Henry Hitchings (2005: 110) notes that:

> . . . dictionaries are fraught with submerged ideas, narratives and histories. Johnson's is no exception. It offers no overarching system of knowledge, but it is a literary anthology, a compendium of quotable nuggets, and a mine of information – some trivial, some considerable – on subjects as diverse as heraldry and hunting, rhetoric and pharmacy, oracles and literary style, the zodiac and magic, law and mathematics, ignorance and politics, the art of conversation and the benefits of reading.

Johnson's dictionary, in other words, contains a lot about Johnson himself – both his interests and his prejudices. It was quite a comprehensive

Some Johnsonian definitions:

urim: *Urim and thummim were something in Aaron's breastplate; but what, criticks and commentators are by no means agreed.*

trolmydames: *[Of this word, I know not the meaning.]*

worm (v.): *To deprive a dog of something, nobody knows what, under his tongue, which is said to prevent him, nobody knows why, from running mad.*

network: *Anything reticulated or decussated, at equal distances, with interstices between the intersections.*

pastern: *The knee of an horse.*

dictionary in its time. But Hitchings notes that Johnson still left out entries for such words as *ultimatum, irritable, zinc, engineering, athlete,* and *annulment*, even though he actually used some of those words in his definitions. On the other hand, he included such words as *ariolation, clancular, deuteroscopy,* and *incompossiblity*, which even the nineteenth-century American lexicographer Noah Webster considered dubious. And it has often been pointed out that some of Johnson's definitions were odd, unhelpful, and occasionally downright wrong.

For example, the word *urim* is used by Milton, and therefore Johnson judged it important enough to be included even though he was unable to discern the meaning of the word from its literary context. Similarly, *trolmydames* is used in Shakespeare, and therefore it merited inclusion for Johnson – although, again, he had no idea what it meant. And what can we say about the definitions for *network* and *worm*? If you don't already know what they mean, you won't be enlightened by Johnson's definitions!

As Hitchings implies in the passage quoted above, we can learn a lot about Johnson's interests from his dictionary. For example, we can tell from Johnson's entry for *pastern* that he had no particular knowledge of or interest in horses: he defines the *pastern* as the knee of a horse. People who are interested in horses know that a *pastern* is part of a horse's foot. Similarly, as Hitchings points out, Johnson apparently had no interest in music. His definitions for a number of stringed instruments (*viola, lute, guitar*) are precisely the same: "a stringed instrument." Furthermore, the definition of *violin* suffers from the cardinal lexicographical sin of circularity: the entry for *violin* sends one to the entry for *fiddle*, which in turn sends one back to *violin*.

These examples are not intended to imply that Johnson's dictionary was incompetent – far from it, it was an amazing achievement for one man working alone for seven years. Much of it still holds up to twenty-first century scrutiny. For every entry that is obscure, weird, or unhelpful, there are a hundred that are brilliant and insightful. I devote this much attention to its deficiencies merely to point out that dictionaries are fallible, and often reflect the foibles of their makers.

2.4.3 Webster's dictionary

Johnson's dictionary was followed in 1828 by Noah Webster's dictionary – billed as the first American dictionary. Webster's agenda in writing his dictionary was at least partly political; through the dictionary he sought to establish American English as a national language. His dictionary included not only new words but also new meanings that had developed for old words in the context of American life, for example, words relevant to the newly minted form of democracy, such as *congress* and *senate*. Webster is also credited with promoting the spelling differences which even today distinguish American from British English – *color* instead of *colour*, *center* instead of *centre*, *tire* instead of *tyre*, and so on.

Webster was not particularly skilled at **etymology** (the study of where words come from); Baugh and Cable (1993: 361) suggest that his sense of

nationalism caused him to ignore advances in historical and comparative linguistics that were taking place in Europe at that time. However, his definitions are excellent. Not surprisingly, though, some definitions in Webster's dictionary are pirated directly from Johnson. Note, however, that not all of Webster's contemporaries shared his desire to distinguish American English from British English. Joseph Worcester, for example, published his own *Comprehensive Pronouncing and Explanatory Dictionary* in 1830, in which he took a far more conservative approach to Americanisms and spelling.

2.4.4 The Oxford English Dictionary

By the mid-nineteenth century, members of the English Philological Society had come to feel that Johnson's dictionary was inadequate. As we saw above, Johnson had missed many words, and even if he had not, over the course of a century many new words are added to a language and many old words come to be used in new ways. After much deliberation and a number of false starts, the Philological Society chose James Murray, a Scottish schoolmaster, to edit the *New English Dictionary*. Oxford University Press contracted to publish it, and by 1895 it had come to be known as the *Oxford English Dictionary*. Murray began work on the dictionary in 1879, hoping to finish it within ten years. But it would be almost fifty years before the first edition of the dictionary was finished, during which time three more editors were added, and Murray himself died.

The OED took so long to compile because the goals of its originators were so ambitious. Murray and his colleagues sought to create a dictionary that would not only give current meanings of words, but also trace those words back as far into the history of English as they could, taking note of all the spelling variants and meaning changes along the way. Following Johnson's dictionary, all senses of words would be illustrated with quotations from literary works. Words that were already archaic or obsolete by the late nineteenth century would still be included, as long as they had not died out before 1250 CE. The dictionary was to be comprehensive in both breadth and depth, a task which turned out to be far more challenging than anyone in 1879 could have anticipated. The first edition of the OED ran to ten large volumes and contained almost a quarter of a million main entries. By the time the last volume was finished, the early volumes were already obsolete; one supplement was added in 1933, and a second one in 1972. A second edition of twenty volumes was issued in 1989, incorporating all of the supplements into the original volume. Today, work continues on the third edition, with segments issued on-line on a quarterly basis, as they are finished. Since the first edition, the OED has grown to include more than half a million entries; in its on-line form, size and space are no longer as much of a concern as they once were.

James Murray was well aware both of the weight his lexicographical decisions carried and of his potential fallibility in making those decisions – after all, most people do look to the dictionary to determine whether *xyz*

Of Obscure Meaning in the OED:

These entries for smazky *and* val-dunk *are among the 87 entries that the OED designates as "meaning obscure" or "of obscure meaning". Did they deserve to be in the OED? You decide!*

smazky, a. *Obs. (Meaning obscure.)*
1599 MIDDLETON Micro-cynicon A5, *Auant,..Ile anger thee inough, And fold thy firy-eyes in thy smazkie snufe.*

val-dunk *Obs. (Meaning obscure.)*
1631 R. BRATHWAIT Whimzies, Wine-soaker *102 By this time his cause is heard, and now this val-dunke growne rampant-drunke, would fight if hee knew how.*

really is a word. Perhaps Murray put it best when he noted in the Introduction to the first edition that:

> The Vocabulary of a widely-diffused and highly-cultivated living language is not a fixed quantity circumscribed by definite limits. That vast aggregate of words and phrases which constitutes the Vocabulary of English-speaking men presents, to the mind that endeavours to grasp it as a definite whole, the aspect of one of those nebulous masses familiar to the astronomer, in which a clear and unmistakable nucleus shades off on all sides, through zones of decreasing brightness to a dim marginal film that seems to end nowhere, but to lose itself imperceptibly in the surrounding darkness.

In other words, it's impossible to pin down the vocabulary of English (and we might add, any other language). Murray illustrates his point with a diagram (figure 2.1), reproduced from the Introduction to the first edition of the dictionary. His idea is that there is a core of words whose place in the dictionary nobody would dispute, encompassing what he called "common," "literary," and "colloquial" words. Common words are words that occur in all registers of English, like *mother, dog, walk, apologetic, wiggle, if, and, to, in, that*, and so on. Literary words are words that we might recognize when we read, but would not necessarily use in daily conversation, words, for example, like *omnipotent, notwithstanding, heretical, avatar,* and *ambulatory*. And also among the core words would be colloquial words, ones that we use frequently in spoken language, but far less frequently in written or formal language, for example, *grubby, pooch,* and *mad* (in the sense of 'angry'). But there is no clear dividing line between these words and words which are perhaps too technical or scientifically specialized (*circumfix, triptan*), not quite assimilated enough into English (*tchachka, sambal oelek*), too bound to a specific dialect (*frappé, black ice*), or too informal, impermanent, or bound too narrowly to a particular time or a particular segment of society (*groovy, homie*). Deciding which of these uncommon words merit inclusion in the dictionary is a judgment call, often based more on practical considerations – the size of the dictionary, its intended audience – than on strict

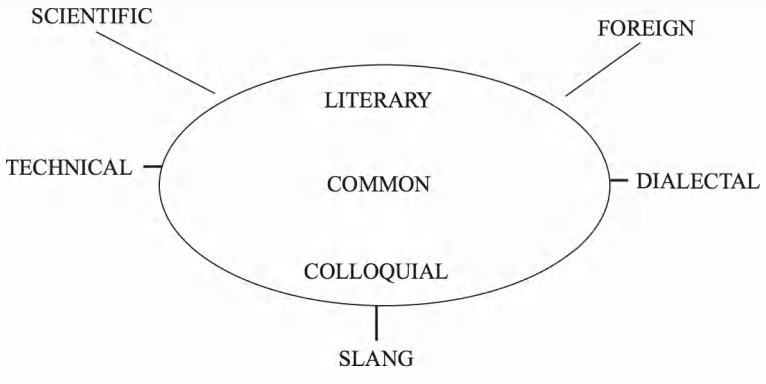


FIGURE 2.1
Reproduced with permission of Oxford University Press, James A.H. Murray et al. 1888. *A New English Dictionary on Historical Principles*, p. xvii

linguistic principles. All lexicographers face this conundrum, and each one makes a slightly different decision.

The OED is certainly the gold standard for English dictionaries today. Nevertheless, it has its own idiosyncrasies. For example, it contains a number of nonce words, words that are attested only once. Indeed there are quite a few nonce words that the OED includes, even though it is unable to define them. A search, using the key words "meaning obscure" and "of obscure meaning," turns up 87 words so labeled, including *smazky, squirgliting, val-dunk, vezon, uncape,* and *umbershoot.* Each bears an entry illustrated with one quotation, which unfortunately does not illuminate the word's meaning. Nevertheless, these and 81 others like them made it into the OED!

2.4.5 Modern dictionaries

Today, there are dozens of dictionaries available for English – unabridged dictionaries, college dictionaries, children's dictionaries, specialized dictionaries of music or architecture, an official Scrabble dictionary, not to mention on-line dictionaries in many varieties. Each one of these is edited by a team of individuals who make the judgment call whether *xyz* deserves to be in the dictionary. The decision is made on a number of grounds:

- the size of the dictionary, which determines the number of words it can hold;
- the intended audience of the dictionary (adults, children, language learners, etc.);
- whether a word has a sufficiently broad base of usage;
- whether it's likely to last;
- whether it's too specialized or technical for the intended audience;
- for a word borrowed from another language, whether it's assimilated enough to be considered part of English.

With respect to size, the number of words and the depth of entries (whether etymologies and illustrative quotes are included, for example) in print dictionaries are determined by the number of pages and the font size of the print used. On-line dictionaries do not have the sort of space constraints that print dictionaries do. As for audience, a dictionary intended for college-age adults will probably have more learned and technical words than a dictionary for children. On the other hand, words that a native speaker is unlikely to need defined might be more of a focus in a dictionary for English language learners; the meanings of prepositions and their idiomatic uses come to mind here. The type of dictionary also determines how broad a base of usage a word needs to have in order to be included. Dictionaries of slang, dialect, or of specialized fields obviously contain more narrowly used words than general dictionaries do (although you might be surprised at how much slang and technical terminology can be found in general dictionaries).

Perhaps the trickiest issue is how long a word has to have been around to merit inclusion in the dictionary. These days, words can appear in

dictionaries fairly quickly, especially in on-line dictionaries. The OED already lists *google* as a verb, with its first illustrative quotation dated 1999. The word *bouncebackability* – allegedly coined by British sportscaster Iain Dowie in 2004 (Hohenhaus 2006) – already had a draft OED entry by June, 2006 (although interestingly, the OED has traced the word as far back as 1961!).

And there is more to consider in deciding whether a word goes into the dictionary. Take, for example, words formed with various prefixes and suffixes. If *happy* is in the dictionary (as it certainly would be), do we need to have an entry as well for *happiness*? Similarly, if *sad* has an entry, do we need *sadness*? If our audience is a learner of English, perhaps yes, but for native speakers who know intuitively how the suffix *-ness* is used, is there any need for these extra entries? Interestingly, dictionaries are often quite inconsistent on how many and which derivatives with particular suffixes get entries. The on-line OED has entries for *redness, blueness, pinkness, greenness,* and *yellowness*, but not *orangeness*. The word *purpleness* is used in the definition of the word *purplely*, but does not have its own entry. And not surprisingly, there is no entry for *mauveness* or *beigeness*. What is more surprising is that there are so many entries for color words with the suffix *-ness* attached.

Certainly, if a word derived with a prefix or suffix takes on an idiosyncratic or **lexicalized** meaning, the dictionary needs to include it. Take, for example, the word *transmission*, which can have the transparent meaning 'the act of transmitting' but probably more often is used to denote a part of a car. This second meaning probably deserves to be in the dictionary. But is it necessary to include all derived words whose meanings are perfectly clear from the meaning of the base plus the meaning of the affix? Probably not.

Until the last decade of the twentieth century lexicographers made their decisions by reading materials of all sorts, and in more recent decades by listening to radio, TV, and talk in general. Potential entries would be recorded with their context on small slips of paper. These slips would then be filed, and when a critical mass of usages accumulated for a word, it might be considered for entry in the dictionary. These days lexicographers are aided by **corpora** (singular **corpus**), large computerized databases that can be searched for words in the context of their use, and by the internet, which might be viewed as a vast corpus. Indeed the rise (and sometimes fall) of a new word can be traced by searching for its use on the internet.

Perhaps the most interesting recent development in lexicography is the rise of Wiktionary – an on-line collaborative dictionary created not by professional lexicographers, but by users themselves. In the instructions for submitting entries, Wiktionary asks that words be attested, by which it means they must be in widespread use, available in well-known works or refereed publications, used at least three times in at least three sources over more than a year. It does, however, have a category of what it calls 'protologisms' for "terms defined in the hopes that they will be used, but which are not actually in wide use."

One final note about the vagaries of dictionaries. Lest you think that lexicographers are humorless ("harmless drudges" as Johnson calls them in his dictionary), let's consider the issue of mountweazels mentioned briefly above. As Henry Alford reveals in the August 29, 2005 issue of *The New Yorker*, the editors of the *New Oxford American Dictionary* (2001) planted the non-existent word *esquivalience* (defined as "the willful avoidance of one's official responsibilities . . .") among the entries for the letter "e" to catch potential dictionary pirates. Such false words are called 'mountweazels', from the false entry for Lillian Virginia Mountweazel in the *New Columbia Encyclopedia*.[3] What is most interesting for our purposes is that once these fake words have been coined, they take on lives of their own. As of December 2006, there were 55,300 hits for *esquivalience* and 22,700 for *mountweazel* on Google, leading me to wonder whether these fakes have now become real words.

2.4.6 And other languages

I have concentrated here on the history of dictionary-making in English, but the same points might be made with respect to dictionaries of French, Italian, Russian, Chinese, or Central Alaskan Yup'ik. All dictionaries are products of individuals and all display the choices and idiosyncrasies of those individuals in some way or another.

Dictionaries of other languages might be organized quite differently from those of the Indo-European languages that we are most familiar with, however. For example, dictionaries of Mandarin Chinese are not alphabetized in the way that dictionaries of English and French are, because Chinese is not written in the Roman alphabet. Instead, the writing system (or **orthography**) of Chinese is **logographic** or word-based. Each word in Chinese is represented by a single character (or sometimes a combination of two characters). When you look up a word in a Chinese dictionary, you need to know how many strokes or lines make up that character. Dictionaries are organized from those characters made up of the fewest strokes to those containing the most strokes.

Dictionaries of other languages might include many fewer complex words than English dictionaries typically do. For example, if a language has very regular rules of word formation such that both the form and the resulting meaning of a complex word are perfectly predictable, the dictionary will have no need to list all complex words in separate entries. All it needs to do is list individual morphemes with their meanings (and perhaps some indication of how they combine). But the less predictable the form and meaning of complex words are, the greater the need to put them in the dictionary.

3. According to the *New Columbia Encyclopedia*, Lillian Virginia Mountweazel lived from 1942 to 1973. A fountain designer and photographer, she was supposedly well known for taking pictures of rural American mailboxes. She died tragically in an explosion while she was on assignment for *Combustibles* magazine.

Summary

In this chapter we have been concerned with the question of what constitutes a word. We have contrasted dictionaries with the mental lexicon. Dictionaries are written constructs that record words, along with their pronunciations, meanings, etymologies, and perhaps examples of use. On the one hand, they do not and cannot contain everything that a native speaker would recognize as words of her language – dictionaries have no need to record regularly inflected forms of words and words derived by very active rules of word formation, for example. On the other hand, dictionaries may include items that perhaps don't deserve to be considered real words – for example, nonce words that are undefinable, or artificially created words put in to check for copyright violations.

Our mental lexicons are something different, however. High-school educated adults may have vocabularies of 60,000 words. We acquire these words rapidly, and sometimes our mental representations are sketchy or incomplete. The evidence we have looked at from aphasia and genetic disorders, as well as studies using PET scans, allows us to begin to develop a picture of how these vast numbers of words are organized in our minds. Unlike dictionaries that list words alphabetically, our mental lexicon is organized as a complex web of entries that are linked in various ways, along with a system of rules for combining listed forms. It appears that entries and rules are at least to some extent wired into different parts of the brain.

Exercises

1. Go to the OED On-line website and search for words that are in the dictionary but have no known definition. To do this, click on Advanced Search (look towards the bottom of the OED home page), and type into the first open box "meaning obscure" or "of obscure meaning." Then choose three words and read through their entries. Do you think the OED was justified in including these words? If so, why? If not, why not?
2. Make a list of five words that you consider to be *slang*. Now look them up in your dictionary (you may use any dictionary at hand, whether print or on-line). First note whether or not you find them. If you do, is the dictionary definition the one that you had in mind? Does your dictionary list them as slang? If not, speculate on why they might not be listed as slang.
3. Make a list of at least ten words that come to mind that end in the suffix *-less*. Look these words up in a dictionary (you may use a standard college desk dictionary like the *American Heritage Dictionary* or you may use the on-line OED). How many of your words are in the dictionary? Is there any pattern that you can discern with respect to the words that are listed, as opposed to the words that are not?
4. Visit the Word Spy website (http://www.wordspy.com). Look at the list of new words and decide which ones, if any, are part of your own mental lexicon. If some of them are, compare your understanding of them with the definition that Word Spy gives.