Cambridge Textbooks in Linguistics

Word-Formation in English

Second Edition

Ingo Plag

Word-Formation in English

Second edition

INGO PLAG

Heinrich-Heine-Universität Düsseldorf

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107172098
DOI: 10.1017/9781316771402

First published 2003
10th printing 2014
Second edition 2018

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall, 2018

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Plag, Ingo author.
Title: Word-formation in English / Ingo Plag.
Description: Second edition. | Cambridge ; New York : Cambridge University Press, 2018. |
Series: Cambridge textbooks in linguistics | Includes bibliographical references and index.
Identifiers: LCCN 2017059711 | ISBN 9781107172098 (hardback : alk. paper) |
ISBN 9781316623299 (pbk. : alk. paper)
Subjects: LCSH: English language–Word formation. | English language–Morphology.
Classification: LCC PE1175 .P55 2018 | DDC 425/.9–dc23
LC record available at https://lccn.loc.gov/2017059711

ISBN 978-1-107-17209-8 Hardback
ISBN 978-1-316-62329-9 Paperback

Contents

Preface to the Second Edition xi
Preface to the First Edition xiii
Abbreviations and Notational Conventions xv

Introduction: What This Book Is about and How It Can Be Used 1

1 Basic Concepts 4
1.1 What Is a Word? 4
1.2 Studying Word-Formation 9
1.3 Inflection and Derivation 13
1.4 Summary 17

2 Studying Complex Words 20
2.1 Identifying Morphemes 20
2.1.1 The Morpheme as the Minimal Linguistic Sign 20
2.1.2 Problems with the Morpheme: The Mapping of Form and Meaning 22
2.2 Allomorphy 27
2.3 Establishing Word-Formation Rules 30
2.4 Multiple Affixation and Compounding 39
2.5 Summary 41

3 Productivity and the Mental Lexicon 44
3.1 Introduction: What Is Productivity? 44
3.2 Possible and Actual Words 45
3.3 Complex Words in the Lexicon 47
3.4 Measuring Productivity 51
3.5 Constraining Productivity 59
3.5.1 Pragmatic Restrictions 60
3.5.2 Structural Restrictions 60
3.5.3 Blocking 63
3.6 Summary 67

4 Affixation 71
4.1 What Is an Affix? 71
4.2 How to Investigate Affixes: More on Methodology 73

4.3 General Properties of English Affixation 77
4.3.1 Phonological Properties 77
4.3.2 Morphological Properties 83
4.3.3 Semantic Properties 84
4.3.4 Classifying Affixes 86
4.4 Suffixes 87
4.4.1 Nominal Suffixes 87
4.4.2 Verbal Suffixes 92
4.4.3 Adjectival Suffixes 94
4.4.4 Adverbial Suffixes 97
4.5 Prefixes 98
4.6 Infixation 100
4.7 Summary 102

5 Derivation without Affixation 105
5.1 Conversion 105
5.1.1 The Directionality of Conversion 106
5.1.2 Conversion or Zero-Affixation? 110
5.1.3 Conversion: Syntactic or Morphological? 112
5.2 Prosodic Morphology 114
5.2.1 Truncations: Truncated Names, -Y Diminutives, and Clippings 115
5.2.2 Blends 120
5.3 Abbreviations and Acronyms 124
5.4 Summary 128

6 Compounding 131
6.1 Recognizing Compounds 131
6.1.1 What Are Compounds Made Of? 131
6.1.2 More on the Structure of Compounds: The Notion of Head 134
6.1.3 Canonical and Non-Canonical Compounds 135
6.1.4 Summary 136
6.2 An Inventory of Compounding Patterns 136
6.3 Nominal Compounds 137
6.3.1 Headedness 138
6.3.2 Interpreting Nominal Compounds 141
6.3.3 Stress Assignment 145
6.4 Adjectival Compounds 150
6.5 Verbal Compounds 151
6.6 Neoclassical Compounds 152
6.7 Compounding: Syntax or Morphology? 156
6.8 Summary 159

7 Theoretical Issues: Modeling Word-Formation 162
7.1 Introduction: Why Theory? 162
7.2 Phonology–Morphology Interaction 163

7.3 Affix Ordering 171
7.4 The Nature of Word-Formation Rules 177
7.4.1 Morpheme-Based Morphology 177
7.4.2 Word-Based Morphology 181
7.4.3 Analogy 186
7.4.4 Naive Discriminative Learning 188
7.5 Summary 194

Answer Key to Exercises 198
References 230
Subject Index 239
Affix Index 242
Author Index 244

Preface to the Second Edition

The first edition of this textbook was published 20 years after Laurie Bauer's *English Word-Formation*, which, I thought, was becoming outdated at the time. After 14 years the same fate had befallen my own book and it was time for an updated second edition. In particular, the work with Laurie Bauer and Shelly Lieber on *The Oxford Reference Guide to English Morphology*, which appeared in 2013, showed me that certain concepts and theoretical notions needed to be reconceived and modernized in the light of the new evidence that had become available by that time. I am very grateful to Laurie and Shelly for the stimulation that originated from our joint work.

Numerous colleagues and students have provided useful comments on the first edition, and I have tried to do my best to incorporate all suggestions into the new edition. In the preparation of the second edition, two anonymous readers for Cambridge University Press provided critical and constructive comments on the book proposal and the first edition. One of them later also provided meticulous and very helpful comments on the first draft of the second edition and spotted numerous errors. Many thanks to all of them. Any remaining errors are, of course, all mine.

The drafts of the manuscript for the second edition have received ample feedback from the researchers in my lab. I am very grateful to Marios Andreou, Sonia Ben Hedia, Heidrun Dorgeloh, Ute Engemann, Christine Günther, Lea Kawaletz, Sven Kotowski, Gero Kunter, Julia Muschalik, Annika Schebesta, Christian Uffmann, and Julia Zimmermann. Sabine Arndt-Lappe gave very useful hints concerning Chapter 7 and many individual points throughout the book. Ricardo Bermúdez-Otero provided much appreciated helpful feedback on the morpho-phonological issues discussed in Chapter 7. Special thanks go to the student assistants who helped to create the best possible manuscript and provided valuable comments from a student perspective: Adiseu Davydova, Olga Dmitrieva, Elena Girardi, Tobias Höhn, Yannick Lauter, Simone Neeten, Mandy Rilke, Viktoria Schneider, Simon Stein, Dila Turus, Lara Wikert. I am also very grateful to Ulrike Kayser, who always keeps the ball rolling for all of us.

This book is for the wonderful people I have on my team.

Abbreviations and Notational Conventions

A	adjective
Adv	adverb
AP	adjectival phrase
C	consonant
n_1	hapaxlegomenon
N	noun *or* number of observations
NDL	Naive Discriminative Learning
NP	noun phrase
P	productivity in the narrow sense
*P**	global productivity
PP	prepositional phrase
PrWd	prosodic word
V	verb *or* vowel
V	extent of use
VP	verb phrase
WFR	word-formation rule
♦	is morphologically related to
~	alternates with
#	word boundary
·	syllable boundary
|	in the context of
< >	orthographic representation
/ /	phonological (i.e. underlying) representation
[ ]	phonetic representation *or* structural boundary
*	impossible word
!	possible, but unattested word
ˊ	main stress
ˋ	secondary stress
σ	syllable
( )	foot boundaries

Introduction: What This Book Is about and How It Can Be Used

The existence of words is usually taken for granted by the speakers of a language. To speak and understand a language means – among many other things – knowing the words of that language. The average speaker knows thousands of words, and new words enter our minds and our language on a daily basis. This book is about words. More specifically, it deals with the internal structure of complex words, i.e. words that are composed of more than one meaningful element. Take, for example, the word *meaningful*, which could be argued to consist of two elements, *meaning* and *-ful*, or even three, *mean*, *-ing*, and *-ful*. We will address the question of how such words are related to other words and how the language allows speakers to create new words. For example, *meaningful* seems to be clearly related to *colorful*, but perhaps less so to *awful* or *plentiful*. And, given that *meaningful* may be paraphrased as 'having (a definite) meaning,' and *colorful* as 'having (bright or many different) colors,' we could ask whether it is also possible to create the word *coffeeful*, meaning 'having coffee.' Under the assumption that language is a rule-governed system, it should be possible to find meaningful answers to such questions.

This area of study is traditionally referred to as 'word-formation,' and the present book is mainly concerned with word-formation in one particular language, English. As a textbook for an undergraduate readership, it presupposes very little or no prior knowledge of linguistics and introduces and explains linguistic terminology and theoretical apparatus as we go along. Important technical terms appear in bold print when first mentioned. Definitions of terms can be easily located via the Subject Index, in which the respective page numbers are given in bold print.

The purpose of the book is to enable students to engage in (and enjoy!) their own analyses of English (or other languages') complex words. After having worked with the book, the reader should be familiar with the necessary and most recent methodological tools to obtain relevant data, such as introspection, electronic text collections (known as 'corpora'), linguistic databases, various types of dictionaries, basic psycholinguistic experiments, and various other resources available on the internet. Furthermore, the reader should be able to systematically analyze their data and to relate their findings to theoretical problems and debates. The book is not written from the perspective of a particular theoretical framework and draws on insights from various research traditions.

Word-Formation in English can be used as a textbook for a course on word-formation (or for the word-formation parts of morphology courses), as a source book for teachers, as a guidebook for student research projects, as a book for self-study by more advanced students (e.g. for their exam preparation), and as an up-to-date reference concerning selected word-formation processes in English for a more general readership.

For each chapter there are a number of basic and more advanced exercises, which are suitable for in-class work or as students' homework assignments. The more advanced exercises include proper research tasks, which also give the students the opportunity to use the different methodological tools introduced in the text. Students can control their learning success by comparing their results with the answer key provided at the end of the book. The answer key features two kinds of answers. Basic exercises always receive definite answers, while for the more advanced tasks sometimes no 'correct' answers are given. Instead, methodological problems and possible lines of analysis are discussed. Even readers not interested in working on the exercises may find it fruitful to read the answer key texts for the advanced exercises, since they broaden and deepen the discussion of certain questions raised in the pertinent chapters.

Those who consult the book as a general reference on English word-formation may check subject, affix, and author indexes and the list of references in order to quickly find what they need. Each chapter is also followed by a list of recommended further reading.

As every reader knows, English is spoken by hundreds of millions of people, and there exist numerous varieties of English around the world. The variety that has been taken as a reference for this book is General American English. The reason for this choice is purely practical: it is the variety the author knows best. Although there are sometimes differences observable between different varieties (especially concerning aspects of pronunciation), with regard to most of the phenomena discussed in this book, different varieties of English show very similar patterns. Mostly for reasons of space, but also due to the lack of pertinent studies, existing differences will not be discussed. However, the book should hopefully enable the readers to adapt and relate the findings presented with reference to American English to the variety of English they are most familiar with.

The structure of the book is as follows. Chapters 1 through 3 introduce the basic notions needed for the study and description of word-internal structure (Chapter 1), the problems that arise with the implementation of these notions in the actual analysis of complex words in English (Chapter 2), and one of the central problems in word-formation, productivity (Chapter 3). The descriptively oriented Chapters 4 through 6 deal with the different kinds of word-formation processes that can be found in English: Chapter 4 discusses affixation, Chapter 5 non-affixational processes, and Chapter 6 compounding. Chapter 7 is devoted to two theoretical issues, the role of phonology in word-formation and the nature of word-formation rules.

The author welcomes comments and feedback on all aspects of this book, especially from students. Without students telling their teachers what is good for them (i.e. for the students), teaching cannot become as effective and enjoyable as it should be for both teachers and teachees (oops, was that a possible word of English?).

But there is another way in which the sound structure can tell us something about the nature of the word as a linguistic unit. Think of stress. In many languages (including English) the word is the unit that is crucial for the occurrence and distribution of stress. Spoken in isolation, every word can have only one **main stress** (also called **primary stress** if there are other stressed syllables in the word), as indicated by the acute accents (′) in the data presented in (3) below (note that we speak of linguistic 'data' when we refer to language examples to be analyzed).

(3)

cárpenter	téxtbook
wáter	análysis
féderal	sýllable
móther	understánd

The main stressed syllable is the syllable which is the most prominent one in a word. Prominence of a syllable is a function of loudness, pitch, and duration. Stressed syllables are pronounced louder, with higher pitch, or with longer duration than the neighboring syllable(s). Longer words often have additional, weaker stresses, commonly named **secondary stresses**, which we ignore here for simplicity's sake. The words in (4) now show that the phonologically defined word is not always identical with the orthographically defined word.

(4)

Bénjamin's
gírlfriend
apártment building

While *apártment building* is two orthographic words, it is only one word in terms of stress behavior. The same holds for other compounds like *trável agency, wéather forecast, spáce shuttle*, etc. We see that in these examples the phonological definition of 'word' comes closer to our intuition of what a word should be.

We have to take into consideration, however, that not all words carry stress. For example, function words like articles or auxiliaries are usually unstressed (*a cár*, *the dóg*, *Máry has a dóg*) or even reduced to a single sound (*Jane's in the garden*, *I'll be there*). Hence, the stress criterion is not readily applicable to function words, and to words that hang on to other words (commonly referred to as clitics, e.g. *'ve*, *'s*, *'ll*).

Let us now consider the integrity criterion, which says that the word is an indivisible unit into which no intervening material may be inserted. If some modificational element is added to a word, it must be done at the edges, but never inside the word. For example, plural endings such as *-s* in *girls*, negative elements such as *un-* in *uncommon,* or endings that create verbs out of adjectives (such as *-ize* in *colonialize*) never occur inside the word they modify but are added either before or after the word; hence, the impossibility of formations such as **gi-s-rl*, **com-un-mon*, **col-ize-onial* (note that the asterisk indicates impossible words, i.e. words that are not formed in accordance with the morphological rules of the language in question).

However, there are some cases in which word integrity is violated. For example, the plural of *son-in-law* is not **son-in-laws* but *sons-in-law*. Under the assumption that *son-in-law* is one word, the plural ending is inserted inside the word and not at the end. Apart from certain compounds, we can find other words that violate the integrity criterion for words. For example, in creations like *abso-bloody-lutely*, the element *bloody* is inserted inside the word, and not, as we would expect, attached at one of the edges. In fact, it is impossible to add *bloody* before or after *absolutely* in order to achieve the same effect. *Absolutely bloody* would mean something completely different, and **bloody absolutely* seems utterly strange and uninterpretable.

We can conclude that there are certain, though marginal, counterexamples to the integrity criterion, but surely these cases should be regarded as the proverbial exceptions that prove the rule.

The semantic definition of 'word' states that a word expresses a unified semantic concept. Although this may be true for most words (even for *son-in-law*, which is ill-behaved with regard to the integrity criterion), it is not sufficient in order to differentiate between words and non-words. The simple reason is that not every unified semantic concept corresponds to one word in a given language. Consider, for example, the smell of fresh rain in a forest in the fall. Certainly a unified concept, but we would not consider *the smell of fresh rain in a forest in the fall* a word. In fact, English simply has no single word for this concept. A similar problem arises with phrases like *the woman who lives next door*. This phrase refers to a particular person and should therefore be considered as something expressing a unified concept. This concept is, however, expressed by more than one word. We learn from these examples that although a word may always express a unified concept, not every unified concept is expressed by one word. Hence, the criterion is not very helpful in distinguishing between words and larger units that are not words. An additional problem arises from the notion of 'unified semantic concept' itself, which seems to be rather vague. For example, does the complicated word *conventionalization* really express a unified concept? If we paraphrase it as 'the act or result of making something conventional,' it is not entirely clear whether this should still be regarded as a 'unified concept.' Before taking the semantic definition of 'word' seriously, it would be necessary to define exactly what 'unified concept' means.

This leaves us with the syntactically oriented criterion of wordhood. Words are usually considered to be syntactic atoms, i.e. the smallest elements syntactic rules can refer to. For example, if we say that auxiliary verbs occupy the initial position in 'yes/no' questions (e.g. *Can you come tomorrow?* vs. *You can come tomorrow*), this is evidence that auxiliary verbs are words. This may not come as a surprise to you, as it is probably this criterion that underlies our intuitions about where to put spaces in writing.

Words as syntactic atoms belong to certain syntactic classes (nouns, verbs, adjectives, prepositions, etc.), which are called **parts of speech**, **word classes**, or **syntactic categories**. The position in which a given word may occur in a

sentence is determined by the syntactic rules of a language. These rules make reference to words and the class they belong to. For example, *the* is said to belong to the class called 'articles,' and there are rules which determine where in a sentence such words, i.e. articles, may occur (usually before nouns and their modifiers, as in *the big house*). We can therefore test whether something is a word by checking whether it belongs to such a word class. If the item in question, for example, follows the rules for nouns, it should be a noun, hence a word. Or consider the fact that only words (and groups of words), but no smaller units, can be moved to a different position in the sentence. For example, in 'yes/no' questions, the auxiliary verb does not occur in its usual position but is moved to the beginning of the sentence (*You **can** read my textbook* vs. ***Can** you read my textbook?*). Hence the auxiliary verb must be a word. Thus syntactic criteria can help to determine the wordhood of a given entity.

To summarize our discussion of the possible definition of 'word,' we can say that, in spite of the intuitive appeal of the notion of 'word,' it is sometimes not easy to decide whether a given string of sounds (or letters) should be regarded as a word or not. In the treatment above, we have concentrated on the discussion of such problematic cases. In most cases, however, the stress criterion, the integrity criterion, and the syntactic criteria lead to sufficiently clear results. The properties of words are summarized in (5):

(5) Properties of words
– words are entities having a part-of-speech specification
– words are syntactic atoms
– words (usually) have one main stress
– words (usually) are indivisible units (no intervening material possible)

Unfortunately, there is yet another problem with the word *word* itself. Thus, even if we have unequivocally decided that a given string is a word, some insecurity remains about what exactly we refer to when we say things like

(6) a. The word *be* occurs twice in the sentence.
b. [ðəwɜːrdbiəkɜːrztwaɪsɪnðəsentəns]

For reasons that will become clear below, the utterance in (6) is given in two forms, in the normal spelling (6a), and in phonetic transcription (6b) (throughout the book I will use phonetic transcriptions as given in the *Longman Dictionary of Contemporary English* (*LDCE*), for the North American English pronunciation).

(6) can be understood in different ways. First, <be> or the sounds [bi] may refer to the letters or the sounds which they stand for. Then sentence (6) would, for example, be true for every written sentence in which the string <BLANK SPACE be BLANK SPACE> occurs twice. Referring to the spoken equivalent of (6a), represented by the phonetic transcription in (6b), (6) would be true for any sentence in which the string of sounds [bi] occurs twice. In this case, [bi] could refer to two different 'words,' e.g. *bee* and *be*. The third possible interpretation is that in (6) we refer to the grammatically specified form *be*, i.e. the infinitive,

imperative, or subjunctive form of the linking verb BE (as in *To be or not to be…*, *Be quiet!*, *Whether they be friend or foe…*, respectively). Such a grammatically specified form is called the **grammatical word** (or **morphosyntactic word**). Under this reading, (6) would be true of any sentence containing two infinitive, two imperative, or two subjunctive forms of *be*, but would not be true of a sentence which contains any of the forms *am*, *is*, *are*, *was*, *were*.

To complicate matters further, even the same form can stand for more than one different grammatical word. Thus, the **word-form** *be* is used for three different grammatical words, expressing subjunctive, infinitive, or imperative, respectively. This brings us to the last possible interpretation, namely that (6) may refer to the linking verb BE in general, as we would find it in a dictionary entry, abstracting away from the different word-forms in which the word BE occurs (*am*, *is*, *are*, *was*, *were*, *be*, *being*, *been*). Under this reading, (6) would be true for any sentence containing any two word-forms of the linking verb, i.e. *am*, *is*, *are*, *was*, *were*, *be*, *being*, and *been*. Under this interpretation, *am*, *is*, *are*, *was*, *were*, *be*, *being*, and *been* are regarded as realizations of an abstract morphological entity. Such abstract entities are called **lexemes**. Coming back to our previous example of *be* and *bee*, we could now say that BE and BEE are two different lexemes that simply sound the same (usually small capitals are used when writing about lexemes). In technical terms, they are **homophonous** words, or simply **homophones**.

In everyday speech, these rather subtle ambiguities in our use of the term 'word' are easily tolerated and are often not even noticed, but when discussing linguistics, it is sometimes necessary to be more explicit about what exactly one talks about. Having discussed what we can mean when we speak of 'words,' we may now turn to the question of what exactly we are dealing with in the study of word-formation.

1.2 Studying Word-Formation

As the term 'word-formation' suggests, we are dealing with the formation of words, but what does that mean? Let us look at a number of words that fall into the domain of word-formation and a number of words that do not:

(7)

a.	b.	c.
employee	apartment building	chair
inventor	greenhouse	neighbor
inability	team manager	matter
meaningless	truck driver	brow
suddenness	blackboard	great
unhappy	son-in-law	promise
decolonialization	pickpocket	discuss

In columns (7a) and (7b), we find words that are obviously composed by putting together smaller elements to form larger words with more complex meanings. We can say that we are dealing with morphologically **complex words**. For

example, *employee* can be analyzed as being composed of the verb *employ* and the ending *-ee*, and the adjective *unhappy* can be analyzed as being derived from the adjective *happy* by the attachment of the element *un-*. And the word *decolonialization* can be segmented into the smallest parts *de-*, *colony*, *-al*, *-ize*, and *-ation*. We can thus decompose complex words into their smallest meaningful units. These units are called **morphemes**.

In contrast to those in (7a) and (7b), the words in (7c) cannot be decomposed into smaller meaningful units, they consist of only one morpheme, they are called 'monomorphemic' or 'simplex.' *Neighbor*, for example, is not composed of *neighb-* and *-or*, although the word looks rather similar to a word such as *inventor*. *Inventor* ('someone who invents (something)') is decomposable into two morphemes, because both *invent-* and *-or* are meaningful elements, whereas neither *neighb-* nor *-or* carry any meaning in *neighbor* (a neighbor is not someone who neighbs, whatever that may be . . .).

As we can see from the complex words in (7a), some morphemes can occur only if attached to some other morpheme(s). Such morphemes are called **bound morphemes**, in contrast to **free morphemes**, which do occur on their own. Some bound morphemes, for example *un-*, must always be attached before the central meaningful element of the word, i.e. the **root**, **stem**, or **base**, whereas other bound morphemes, such as *-ity, -ness*, or *-less*, must follow the base. Using Latin-influenced terminology, *un-* is called a **prefix**, *-ity* a **suffix**, with **affix** being the cover term for all bound morphemes that attach to roots. Note that there are also **bound roots**, i.e. roots that occur only in combination with some other bound morpheme. Examples of bound roots are often of Latin origin, e.g. *circul-* (as in *circulate*, *circulation*, *circulatory*, *circular*), *approb-* (as in *approbate*, *approbation*, *approbatory*, *approbator*), or *simul-* (as in *simulant*, *simulate*, *simulation*), but occasional native bound roots can also be found (e.g. *hap-*, as in *hapless*).

Before we turn to the application of the terms introduced in this section, we should perhaps clarify the distinction between 'root,' 'stem,' and 'base,' because these terms are not always clearly defined in the morphological literature and are therefore a potential source of confusion. One reason for this lamentable lack of clarity is that languages differ remarkably in their morphological make-up, so that different terminologies reflect different organizational principles in the different languages. The part of a word which an affix is attached to is called **base**. We will use the term **root** to refer to bases that cannot be analyzed further into morphemes. The term 'stem' is usually used for bases of inflections, and occasionally also for bases of derivational affixes. To avoid terminological confusion, we will avoid the use of the term 'stem' altogether and speak of 'roots' and 'bases' only.

The term 'root' is used when we want to explicitly refer to the indivisible central part of a complex word. In all other cases, where the status of a form as indivisible or not is not at issue, we can just speak of **bases** (or, if the base is a word, of base words). The derived word is often referred to as a **derivative**. The base of the suffix *-al* in the derivative *colonial* is *colony*, the base of the

Sometimes truncation and affixation can occur together, as with formations expressing intimacy or smallness. Such formations are called **diminutives**:

(12) Mandy (♦ Amanda)
Andy (♦ Andrew)
Charlie (♦ Charles)
Patty (♦ Patricia)
Robbie (♦ Roberta)

We also find amalgamations of parts of different words, such as *smog* (♦ ***sm****oke/****fog***) or *modem* (♦ ***mo****dulator/****dem****odulator*), which are called **blends**. The classification of blending either as a special case of compounding or as a case of non-affixational derivation is not so clear. We will return to this issue in section 5.2.2.

Words can also be coined on the basis of individual letters. This can be done in two ways, the first being the combination of the initial letters of compounds or phrases into a word that is pronounced in the usual way, such as *NATO* [neɪtoʊ] or *UNESCO* [juneskoʊ]. Such words are called **acronyms**. The second option are simple **abbreviations** like *UK* or *USA*, in which each letter is individually pronounced.

In sum, there is a host of possibilities speakers of a language have at their disposal (or had so in the past, when the words were first coined) to create new words on the basis of existing ones, including the addition and subtraction of phonetic (or orthographic) material. The study of word-formation can thus be defined as the study of the ways in which new complex words are built on the basis of other words or morphemes. Some consequences of such a definition will be discussed in the next section.

1.3 Inflection and Derivation

The definition of 'word-formation' in the previous paragraph raises an important problem. Consider the italicized words in (13) and think about the question whether *kicks* in (13a), *drinking* in (13b), or *students* in (13c) should be regarded as 'new words' in the sense of our definition.

(13) a. She *kicks* the ball.
b. The baby is not *drinking* her milk.
c. The *students* are not interested in physics.

The italicized words in (13) are certainly complex words, all of them are made up of two morphemes. *Kicks* consists of the verb *kick* and the third-person-singular suffix *-s*, *drinking* consists of the verb *drink* and the participial suffix *-ing*, and *students* consists of the noun *student* and the plural suffix *-s*. However, we would not want to consider these complex words 'new' in the same sense as we would consider *kicker* a new word derived from the verb *kick*.

Here the distinction between word-form and lexeme is again useful. We would want to say that suffixes like participial *-ing*, plural *-s*, or third-person-singular *-s* create new word-forms, i.e. grammatical words, but they do not create new lexemes. In contrast, suffixes like *-er* and *-ee* (both attached to verbs, as in *kicker* and *employee*), or prefixes like *re-* or *un-* (as in *rephrase* or *unconvincing*) do form new lexemes. On the basis of this criterion (i.e. lexeme formation), a distinction has traditionally been made between **inflection** (i.e. called 'conjugation' and 'declension' in traditional grammar) as part of the grammar on the one hand, and **derivation** and compounding as part of word-formation (or rather: lexeme formation).

Let us have a look at the following data which show further characteristics by which the two classes of morphological processes, inflection vs. word-formation, can be distinguished. The derivational processes are in (14a), the inflectional ones in (14b).

(14) a. **derivation**

work*er*	use*less*	*un*truth*fulness*
*inter*view	curios*ity*	passiv*ize*
terror*ism*	*dis*able	econom*ic*

b. **inflection**

(she) work*s*	(the) worker*s*	(is) colonializ*ing*
(we) pick*ed*	(the) child*ren*	John'*s* (house)
Emily'*s* (job)	proud*er*	(has) giv*en*

As already indicated above, the most crucial difference is that inflectional morphemes encode grammatical categories such as plural (*workers*), person (*works*), tense (*picked*), or case (*John's*). These categories are relevant for the building of sentences and are referred to by the grammar. For example, there is a grammatical rule in English that demands that a third-person-singular subject is followed by a verb that is also marked as third-person-singular. This is called subject-verb agreement, which is also relevant for plural marking in sentences (*The flowers are/*is wonderful*). The plural and person suffixes are therefore syntactically relevant, hence inflectional.

One might argue that the suffix *-er* in *worker* is also syntactically relevant, in the sense that it is important for the syntax whether a word is a noun or a verb. That is of course true, but only in a very limited way. Thus, it is not relevant for the syntax whether the noun ends in *-er*, *-ee*, *-ion*, or whether the noun is morphologically complex at all. In that sense, derivational suffixes are not relevant for the syntax.

Let us turn to the next set of properties that unites the words in (14a) and differentiates them from the words in (14b). These properties concern the position of the morphemes: in English derivational morphemes can occur at either end of the base words, whereas regular inflection is always expressed by

suffixes. Only irregular inflection makes use of non-affixational means, as, for example, in *mouse* ♦ *mice* or *sing* ♦ *sang*. There is no inflectional prefix in English. Furthermore, forms like *workers* or *colonializing* indicate that inflectional morphemes always occur outside derivational morphemes; they close the word for further (derivational) affixation (**workers-hood*, **colonializing-er*). As evidenced by derivatives like *un-tru-th-ful-ness* or the famous textbook example *dis-establish-ment-arian-ism*, derivational suffixes can and do occur inside other derivational suffixes.

Another interesting difference between the words in (14a) and (14b) concerns the part of speech. The suffixes in (14a) change the part of speech of the base word. For instance, the suffixation of *-less* makes an adjective out of a noun, the suffix *-ity* makes a noun out of an adjective, and the suffix *-ize* turns an adjective into a verb. The inflectional suffixes don't change the category of the base word. A plural marker on a noun does not change the category, nor does the past-tense marker on the verb. However, not all derivational affixes are category-changing, as is evidenced, for example, by most prefixes (as, for example, in *dishonest*, *decolonialize*, *non-issue*), or by the nominal suffix *-ism*, which can attach to nouns to form nouns (e.g. *terrorism*).

The final property of derivation to be discussed here is exemplified by the two derivatives *interview* and *curiosity* in (14a), as against all inflectional forms. Both forms in (14a) show a property which is often found in derivation, but hardly ever in inflection, and which is called **semantic opacity**. If you consider the meaning of *interview* and the meaning of the ingredient morphemes *inter-* and *view*, you can observe that the meaning of *interview* is not the sum of the meaning of its parts. The meaning of *inter-* can be paraphrased as 'between,' that of (the verb) *view* as 'look at something' (definitions according to the *LDCE*), whereas the meaning of (the verb) *interview* is 'to ask someone questions, especially in a formal meeting.' Thus, the meaning of the derived word cannot be inferred on the basis of its constituent morphemes; it is to some extent **opaque**, or **non-transparent**. The same holds for *curiosity*, a noun that has two related meanings: it can refer to a personal attribute 'the desire to know or learn about anything,' which is transparent, but it can also mean 'object of interest' (cf., for example, the definitions given in the *Oxford English Dictionary*, *OED*), which is certainly less transparent. Non-transparent formations are quite common in derivational morphology, but rare in inflection.

Closely related to this generalization is the fact that inflectional categories tend to be fully **productive**, whereas derivational categories often show strong restrictions as to the kinds of possible combinations. What does 'fully productive' mean? A productive morpheme is one that can be attached regularly to any word of the appropriate class. For example, the morpheme expressing past tense can occur on all regular main verbs. And the morpheme expressing plural on nouns can be said to be fully productive, too, because all count nouns can take plural endings in English (some of these endings are irregular, as in

ox-en, but the fact remains that plural morphology as such is fully productive). Note that the 'appropriate class' here is the class of count nouns; non-count nouns (such as *rice* and *milk*) regularly do not take plural. In contrast to the inflectional verbal and nominal endings just mentioned, not all verbs take the adjectival suffix *-ive*, nor do all count nouns take, say, the adjectival suffix *-al*:

(15)

a.	*walk-ive	exhaust ♦ exhaustive
	*read-ive	operate ♦ operative
	*surprise-ive	assault ♦ assaultive
b.	*computer-al	colony ♦ colonial
	*desk-al	department ♦ departmental
	*child-al	phrase ♦ phrasal

The nature of the restrictions that are responsible for the impossibility of the asterisked examples in (15) (and in derivational morphology in general) are not always clear but are often a complex mixture of phonological, morphological, and semantic mechanisms. The point is that, no matter what these restrictions in derivational morphology turn out to be, inflectional domains usually lack such complex restrictions.

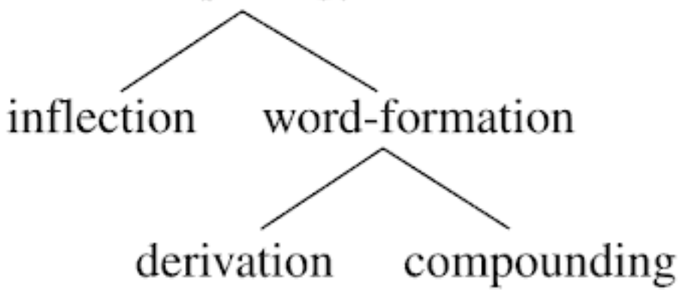
As a conclusion to our discussion of derivation and inflection, I have summarized the differences between inflection and derivation in (16). Exercise 1.6 below focuses on the problems in the application of these criteria and on the general nature of the dichotomy of inflection and derivation:

(16)

derivation	**inflection**
– encodes lexical meaning	– encodes grammatical categories
– is not syntactically relevant	– is syntactically relevant
– can occur inside derivation	– occurs outside all derivation
– often changes the part of speech	– does not change part of speech
– is often semantically opaque	– is rarely semantically opaque
– is often restricted in its productivity	– is fully productive
– is not restricted to suffixation	– always suffixational (in English)

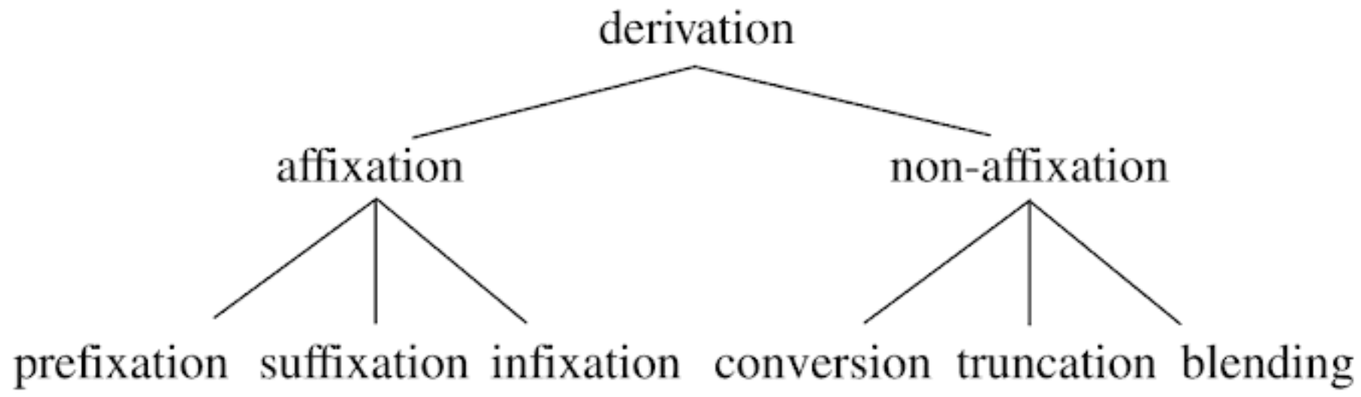
Based on these considerations, we can conclude this subsection by schematically conceptualizing the realm of morphology, as described so far:

(17)



The formal means employed in derivational morphology and discussed so far can be classified in the following way:

(18)



1.4 Summary

In this chapter we have looked at some fundamental properties of words and the notion of 'word' itself. We have seen that words can be composed of smaller units, called morphemes, and that there are many different ways to create new words from existing ones by affixational, non-affixational, and compounding processes. Furthermore, it became clear that there are remarkable differences between different types of morphological processes, which has led us to postulate the distinction between inflection and word-formation.

We are now equipped with the most basic notions necessary for the study of complex words and can turn to the investigation of more (and more complicated) data in order to gain a deeper understanding of these notions. This will be done in the next chapter.

Further Reading

Introductions to the basics of morphological analysis can also be found in other textbooks, such as Haspelmath and Sims (2010), Aronoff and Fudeman (2011), Booij (2012), Fábregas and Scalise (2012), and Lieber (2015). All of these contain useful discussions of the notion of 'word' and introduce basic terminology needed for the study of word-formation. Stockwell and Minkova (2001) and Harley (2006) are introductions to English words, including chapters on morphology. Schmid (2011) and Don (2014) are textbooks focusing on English morphology. Bauer et al. (2013) is a comprehensive and detailed overview of all aspects of English morphology; Plag (2016) is a concise overview of English word-formation at article length. There are also numerous handbooks available that contain state-of-the-art articles on word-formation across many different languages including English: Štekauer and Lieber (2005) and Müller et al. (2016) on word-formation, Lieber and Štekauer (2009) on compounding, Lieber and Štekauer (2014) on derivation.

Those interested in a more detailed treatment of the distinction between inflection and derivation may wish to consult the classic article by Booij (1993), or some of the papers on this topic in one of the handbooks just

2 Studying Complex Words

Outline

This chapter discusses in some detail the problems that arise with the implementation of the basic notions introduced in Chapter 1 in the actual analysis of word structure in English. First, the notion of the morpheme is scrutinized with its problems of the mapping of form and meaning. Then the phenomenon of base and affix allomorphy is introduced, followed by a discussion of the notion of 'word-formation rule.' Finally, cases of multiple affixation and compounding are analyzed.

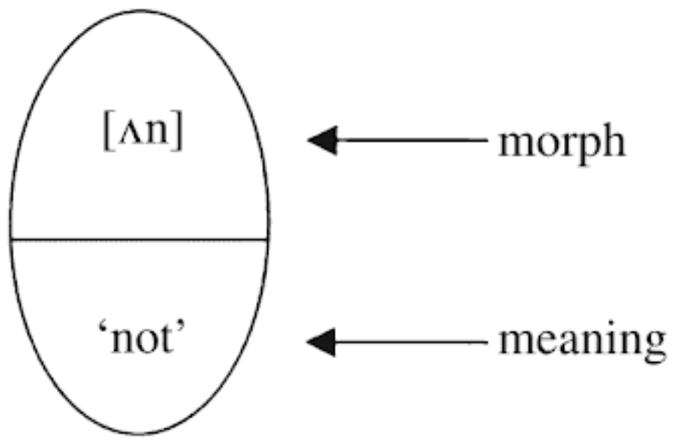
2.1 Identifying Morphemes

In the previous chapter we introduced the crucial notion of the morpheme as the smallest meaningful unit. We have seen that this notion is very useful in accounting for the internal structure of many complex words (recall our examples *employ-ee*, *invent-or*, *un-happy*, etc.). In this section, we will look at more data and see that there are a number of problems involved with the morpheme as the central morphological unit.

2.1.1 The Morpheme as the Minimal Linguistic Sign

The most important characteristic of the traditional morpheme is that it is conceived of as a unit of form and meaning. For example, the morpheme *un-* (as in *unhappy*) is an entity that consists of the content or meaning on the one hand, and the sounds or letters which express this meaning on the other hand. It is a unit of form and meaning, a **sign**. The notion of sign may be familiar to most readers from non-linguistic contexts. A red traffic light, for instance, is also a kind of sign in the above sense: it has a meaning ('stop!'), and it has a form which expresses this meaning. In the case of the traffic light, we could say that the form consists of the well-known shape of the traffic light (a simple torch with a red bulb would not be recognized as a traffic light) and, of course, the red light it emits. Similarly, morphemes have a meaning that is expressed in the physical form of sound waves (in speech) or by the black marks on paper which we call letters. In the case of the prefix *un-*, the unit of form and meaning can be schematically represented as in (1). The part of the morpheme we have referred to as its 'form' is also called **morph**, a term coined on the basis of the Greek word for 'form, figure.'

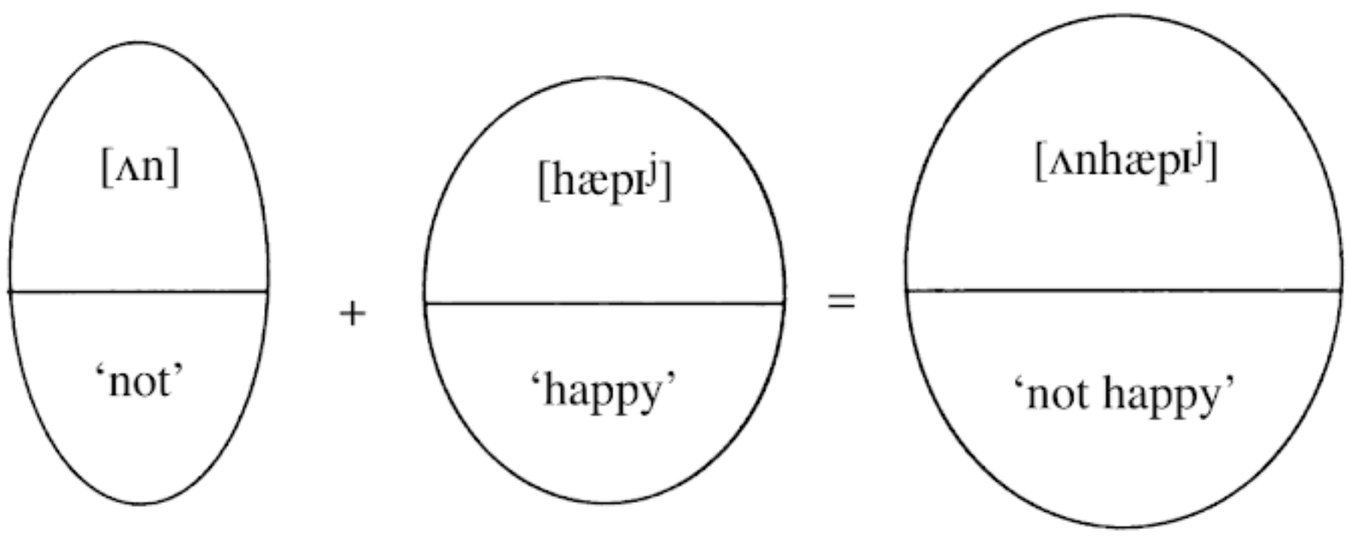
(1) The morpheme *un-*



The pairing of certain sounds with certain meanings is essentially arbitrary. That the sound sequence [ʌn] stands for the meaning 'not' is a pure convention of English, and in a different language (and speech community) the same string of sounds may represent another meaning or no meaning at all.

In complex words at least one morpheme is combined with another morpheme. This creates a derived word, a new complex sign, which stands for the combined meaning of the two morphemes involved. This is schematically shown in (2):

(2)



The meaning of the new complex sign *unhappy* can be predicted from the meanings of its parts. Linguistic expressions such as *unhappy*, whose meaning is a function of the meaning of its parts, are called **compositional**. Not all complex words and expressions, however, are compositional, as can be seen from idiomatic expressions such as *kick the bucket* 'die.' And pairs such as *view* and *interview*, or *late* and *lately*, show that not even all complex words have compositional, i.e. completely transparent, meanings. As we have already seen in the previous chapter, the meaning of the prefix *inter-* can be paraphrased as 'between,' but the verb *interview* does not mean 'view between' but something like 'have a (formal) conversation.' And while *late* means 'after the due time,' the adverb *lately* does not have the compositional meaning 'in a late manner' but is best paraphrased as 'recently.'

A note is in order on what we mean by 'form' if we say that a morpheme is a unit of form and meaning. We can distinguish between two basic kinds of form, acoustic/auditory and visual. In speech we pronounce and hear morphemes or words; hence we are dealing with phonetic or phonological forms that represent an acoustic signal, i.e. our speech. When we write we use letters to represent morphemes or words. The relationship between phonological and orthographic

representations is sometimes quite complex, and it can be observed that sometimes the spelling of a particular morpheme changes if it combines with another morpheme. In the previous chapter, for example, we discussed the derivatives *decolonialization* and *untruthfulness*. You might have noticed that *decolonialization* involved two orthographic adjustments, the change from <y> to <i> in the base *colony*, and the loss of <e> with the suffix *-ize*. The adjective *true* also lost its final <e> under the attachment of *-th*. In this book, I will say very little about spelling alternations and refer the reader instead to Chapter 4 of Bauer et al. (2013), which gives a comprehensive overview of how morphology interacts with spelling.

2.1.2 Problems with the Morpheme: The Mapping of Form and Meaning

One of the central problems with the morpheme is that not all morphological phenomena can be accounted for by a neat one-to-one mapping of form and meaning. Of the many cases that could be mentioned here and that are discussed in the linguistic literature, I will discuss some that are especially relevant to English word-formation.

The first phenomenon which appears somewhat problematic for our notion of the morpheme is conversion, the process by which words are derived from other words without any formal marking (to *walk* ♦ *a walk*, *to throw* ♦ *a throw*, *water* ♦ *to water*, *book* ♦ *to book*). This would force us to recognize morphemes which have *no* morph, which is impossible according to our basic definition of morpheme. We have, however, already seen that this problem can be solved by assuming that zero-forms are also possible elements in language. In this view, the verb *water* is derived from the noun *water* by adding to the base noun *water* a zero-form with the meaning 'apply X.' Thus we could speak of the presence of a **zero-morph** in the case of conversion (hence the competing terms **zero-derivation** or 'zero-suffixation' for conversion). Note that it would be misleading to talk about a 'zero-*morpheme*,' unless you mean by it a morpheme that has a meaning but no form that expresses its meaning.

More serious problems for the morpheme arise when we reconsider the non-affixational processes mentioned in the previous chapter. While affixational processes usually make it easy to find the different morphemes and determine their meaning and form, non-affixational processes generally do not lend themselves to a straightforward analysis in terms of morphemes. Recall that we found a set of words that are derived from other words by truncation (e.g. *Ron*, *Liz*, *lab*, *demo*). Such derivatives pose the question of what exactly the morph is (and where it is) that – together with the base word – forms the derived word in a compositional manner. Perhaps the most natural way to account for truncation would be to say that it is the process of deleting material itself which is the morph. Under this analysis we would have to considerably extend our definition of morpheme ('smallest meaningful element') to allow processes of deletion to

be counted as 'elements' in the sense of the definition. Additionally, the question may arise of what meaning is associated with truncations. What exactly is the semantic difference between *Ronald* and *Ron*, *laboratory* and *lab*? Although maybe not particularly obvious, it seems that the truncations, in addition to the meaning of the base, signal the familiarity of the speaker with the entity s/he is referring to. The marking of familiarity can be seen as the expression of a type of social meaning through which speakers signal their belonging to a certain group. In sum, truncations can be assigned a meaning, but the nature of the morph expressing that meaning is problematic.

In order to save the idea of morphemes as 'things' (instead of processes), one could also propose a different analysis of truncation, assuming the existence of a truncation morpheme which has no phonetic content but which crucially triggers the deletion of phonetic material in the base. Alternatively, we could conceptualize the formal side of the truncation morpheme as an empty morph which is filled with material from the base word.

A similar problem for the morpheme-is-a-thing view emerges from cases like the two verbs *to fall* 'move downwards' and *to fell* 'make fall.' It could be argued that *fell* is derived from *fall* by the addition of a **causative** morpheme 'make X.' This idea is not far-fetched, given that the formation of causative verbs is quite common in English, but usually involves affixes, such as *-ify* in *humidify* 'make humid,' or *-en* in *blacken* 'make black.' But where is the causative morpheme in *to fell*? Obviously, the causative meaning is expressed merely by the vowel change in *fall* vs. *fell* ([ɔ] → [ɛ]) and not by any affix. A similar kind of process, i.e. the addition of meaning by means of **vowel alternation**, is evidenced in English in certain cases of past-tense formation and of plural marking on nouns, as illustrated in (3):

(3)

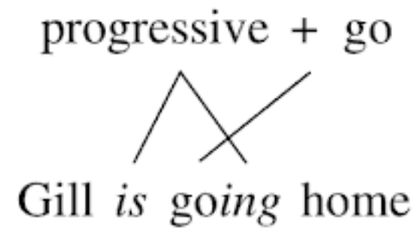
a.		b.	
stick ♦ stuck		foot ♦ feet	
sing ♦ sang		goose ♦ geese	
take ♦ took		mouse ♦ mice	

Again, this is a problem for those who believe in morphemes as elements. And again, a redefinition in terms of processes can save the morpheme as a morphological entity but seriously weakens the idea that the morpheme is a minimal sign, given that signs are not processes, but physical entities signifying meaning.

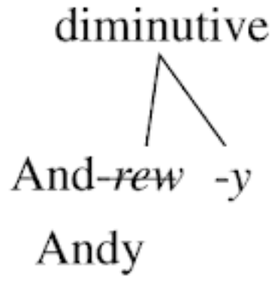
Another problem of the morpheme is that in some expressions there is more than one form signifying a certain meaning. A standard example from inflectional morphology is the progressive form in English, which is expressed by the combination of the verbal suffix *-ing* and the auxiliary verb BE preceding the suffixed verb form. A similar situation holds for English *-y* diminutives, which are marked by a combination of truncation and suffixation, i.e. the absence of parts of the base word on the one hand and the presence of the suffix *-y* on the other hand. Such phenomena are instances of what is known as **extended exponence** because the forms that represent the morpheme extend across more than one element. Extended exponence is schematically illustrated in (4):

(4)

a. progressive in English



b. *-y* diminutives in English



To account for cases of extended exponence, we have to allow morphemes to be **discontinuous**. In other words, we have to allow for the meaning of a morpheme to be realized by more than one morph, e.g. by a form of BE and *-ing* in the case of the progressive, and by truncation and *-y* in the case of diminutives.

Another oft-cited problem of the morpheme is that there are frequently parts of words that invite morphological segmentation, but do not carry any meaning and hence do not qualify for morpheme status. Consider, for example, the following words, and try to determine the morphemes which the words may be composed of:

(5) infer confer prefer refer transfer

A first step in the analysis of the data in (5) may be to hypothesize the existence of a morpheme *-fer* (a bound root) with a number of different prefixes (*in-*, *con-*, *pre-*, *re-*, *trans-*). However, if *-fer* is a bound root, it should have the same (or at least sufficiently similar) meanings in all the words in which it occurs. If you check the meanings these words have in contemporary English in a dictionary, you may end up with paraphrases similar to those found in the *OED*:

(6)

infer	'to draw a conclusion'
confer	'to converse, talk together'
prefer	'to like better'
refer	'to send or direct (one) to a person, a book . . . for information'
transfer	'to convey or take from one place, person, etc. to another'

Those readers who know some Latin may come up with the hypothesis that the words are borrowed from Latin (maybe through French) and that therefore *-fer* means 'carry,' which is the meaning of the Latin root. This works for *transfer*, which can be analyzed as consisting of the prefix *trans-* 'across' and the bound root *-fer* 'carry.' *Transfer* has then the compositional meaning 'carry across, carry over,' which is more or less the same as what we find in the *OED*. Unfortunately, this does not work for the other words in (5). If we assume that *in-* is a prefix meaning 'in, into,' we would predict that *infer* would mean 'carry into,' which is not even close to the real meaning of *infer*. The meaning of *con-* in

More recently, however, a gradient view of morphological complexity has been proposed (see, e.g., Hay and Baayen 2005), according to which words may be more or less complex. In this perspective a morphological boundary may be stronger or weaker. Thus there may be a continuum of boundary strength even among words that have the same suffix. Consider, for example, the word *government*, which seems to be composed of two morphemes, the verb *govern* and the noun-deriving suffix *-ment*, which usually denotes an action or a result, as in *assessment* or *improvement*. There is, however, something peculiar about the word *government*. It is usually pronounced [gʌvmənt]; the base loses its final sounds. Furthermore, there is some degree of semantic opacity: *government* does not mean 'action or result of governing' (as can be expected from a well-behaved *-ment* word) but, rather, refers to the people who govern, or, more generally, to the political authorities. In other words, *government* shows properties that seem to some degree independent of its constituent morphemes. It can thus be argued that *government* is morphologically less complex than, for example, *assessment* or *improvement*, whose phonological and semantic behavior is fully predictable from the morphemes that make up these words. In some words, the fusion of base and affix is so strong that one can even start doubting whether there is still a suffix involved, as in *business* 'company, commerce,' or *listless* 'tired and not interested.' If morphological gradience is assumed to be a general phenomenon, this questions the idea of the morpheme as a clearly separable and discrete unit of form and meaning.

To summarize our discussion of the morpheme so far, we have seen that it is a useful unit in the analysis of complex words, but not without theoretical problems. These problems can, however, be solved in various ways by redefining the morpheme appropriately. For the purposes of this book, it is not necessary to adhere to any particular theory of the morpheme. In most cases morpheme status is uncontroversial, and in controversial cases we will use more neutral terminology. In section 7.4 we will return to the theoretical issues touched upon above.

2.2 Allomorphy

So far we have tacitly assumed that morphemes have invariable realizations. That is, we have assumed that one meaning is expressed by a certain morph or a certain string of morphs and not by variable morphs whose exact shape differs according to the context in which they occur. The above discussion of *government* has shown, however, that this view is probably too simple. With many morphemes, we actually find some variability in their form. For instance, the definite and indefinite articles in English take on different shapes, depending on the kind of word which they precede:

(9) The shape of articles in English

a. the indefinite article *a*

[ə] question	[ən] answer
[ə] book	[ən] author
[ə] fence	[ən] idea

in isolation: ['eɪ]

b. the definite article *the*

[ðə] question	[ði] answer
[ðə] book	[ði] author
[ðə] fence	[ði] idea

in isolation: ['ði]

The data clearly show that there are three distinct realizations of the indefinite article and three distinct realizations of the definite article. When not spoken in isolation, the indefinite article *a* has two different morphs [ə] and [ən], and the definite article *the* equally has two morphs, [ðə] and [ði]. When spoken in isolation (or sometimes when speakers hesitate, as in *I saw a . . . a . . . a unicorn*), each article has a third, stressed, variant, ['eɪ] and ['ði] respectively. Such different morphs representing the same morpheme are called **allomorphs**, and when different morphs realize one and the same morpheme the phenomenon is known as **allomorphy**.

How do speakers know when to use which allomorph? In the case of the articles, the answer is fairly straightforward. One of the two allomorphs occurs when a consonant follows, the other when a vowel follows. The third allomorph occurs if nothing follows. On a more abstract level, we can say that it is the sound structure that **conditions** the **distribution** of the allomorphs, i.e. determines which allomorph has to be used in a given linguistic context. This is called **phonological conditioning**. We will shortly see that there are also other kinds of conditioning factors involved in allomorphy.

Allomorphy is also rather frequent in English derivation, and both bases and affixes can be affected by it. Consider first a few cases of base allomorphy and try to determine how many allomorphs the lexemes EXPLAIN, MAINTAIN, COURAGE have:

(10)

explain	maintain	courage
explanation	maintenance	courageous
explanatory		

To make things more transparent, let us look at the actual pronunciations, given in phonetic transcription in (11) below.

(11)

[ɪkˈspleɪn]	[meɪnˈteɪn, mənˈteɪn]	[ˈkɜːrɪdʒ]
[ˌɛkspləˈneɪʃən]	[ˈmeɪntənəns]	[kəˈreɪdʒəs]
[ɪkˈsplænətɔːrɪ]		

Let us first describe the allomorphy of the bases in (10) and (11). Obviously, the pronunciation of the base EXPLAIN varies according to the kind of suffix attached