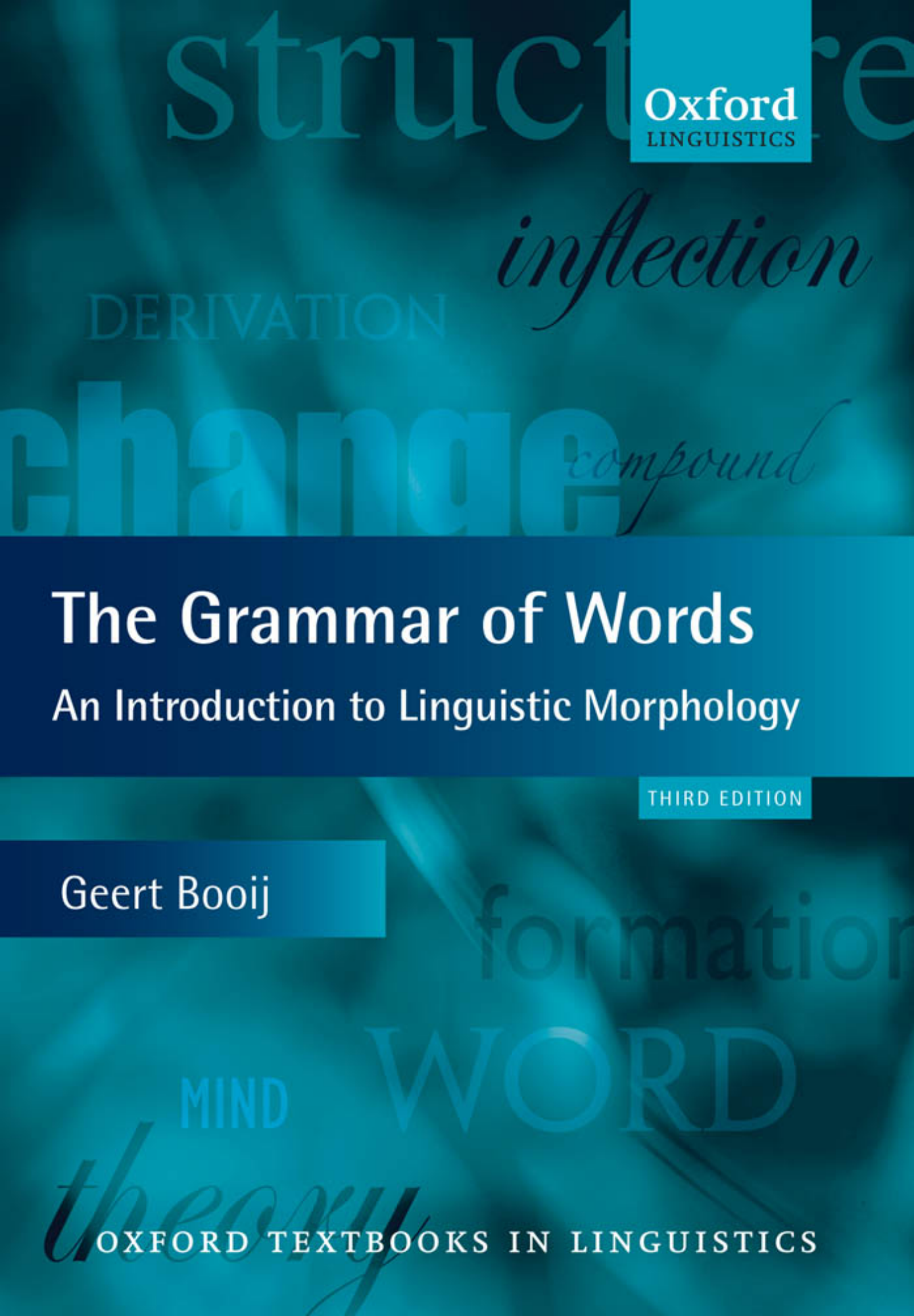
Oxford
LINGUISTICS
inflection
DERIVATION
compound
The Grammar of Words
An Introduction to Linguistic Morphology
THIRD EDITION
Geert Booij
formation
MIND
WORD
theory
OXFORD TEXTBOOKS IN LINGUISTICS

OXFORD TEXTBOOKS IN LINGUISTICS

The Grammar of Words
Third Edition

Oxford Textbooks in Linguistics

PUBLISHED

The Grammar of Words
An Introduction to Linguistic Morphology
Third edition
by Geert Booij

A Practical Introduction to Phonetics
Second edition
by J. C. Catford

Meaning in Use
An Introduction to Semantics and Pragmatics
Third edition
by Alan Cruse

Natural Language Syntax
by Peter W. Culicover

Principles and Parameters
An Introduction to Syntactic Theory
by Peter W. Culicover

A Semantic Approach to English Grammar
by R. M. W. Dixon

Semantic Analysis
A Practical Introduction
by Cliff Goddard

Pragmatics
by Yan Huang

The History of Languages
An Introduction
by Tore Janson

Diachronic Syntax
by Ian Roberts

Cognitive Grammar
An Introduction
by John R. Taylor

Linguistic Categorization
Third edition
by John R. Taylor

IN PREPARATION

Compositional Semantics
An Introduction to the Syntax/Semantics Interface
Pauline Jacobson

The Lexicon
An Introduction
by Elisabetta Ježek

Functional Discourse Analysis
by Evelien Keizer

Semantics and Pragmatics
Meaning in Language and Discourse
Second edition
by Kasia M. Jaszczolt

Translation
Theory and Practice
by Kirsten Malmkjaer

Linguistic Typology
Theory, Method, Data
by Jae Jung Song

The Grammar of Words

An Introduction to Linguistic Morphology

Third edition

Geert Booij

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First edition published in 2005
Second edition published 2007
Third edition published in 2012

Impression: 1

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

ISBN 978–0–19–969183–8

Printed in Great Britain by
Ashford Colour Press Ltd, Gosport, Hampshire

Contents

Typographic Conventions

Small capitals	For lexemes, semantic components, and morphosyntactic and morphosemantic categories
Bold type	For terms first introduced and explained in the Glossary
Italics	For citation forms when not set on different lines
Single quotation marks	For quotations from other authors
Double quotation marks	For glosses
Questions marks	For formal or semantic oddness
Asterisks	For ungrammaticality
//	For underlying phonological representations
[ ]	For phonetic representations

Abbreviations and Symbols

A	Adjective, transitive subject
ABL	Ablative
ABS	Absolutive
ACC	Accusative
AFF,aff	Affix
ALL	Allative
ANTIPASS	Antipassive
AP	Adjective Phrase
APPL	Applicative
AUX	Auxiliary
CAUS	Causative
COM	Comitative
COMP	Completive aspect
COORD	Coordination
CVB	Converb
DAT	Dative
DO	Direct Object
DS	Different Subject
DIM	Diminutive
ERG	Ergative
EVID	Evidential
F	Foot
F, FEM	Feminine
FUT	Future
GEN	Genitive
H	High Tone
IMP	Imperative
IMPF	Imperfective
INCOMP	Incompletive aspect
IND	Indicative
INF	Infinitive
INSTR	Instrumental
INTR	Intransitive

IO	Indirect Object
L	Low Tone
LCS	Lexical Conceptual Structure
LOC	Locative
M	Mid Tone
M, MASC	Masculine
N	Noun, Number of tokens
n_1	number of hapaxes
NEG	Negation
N, NEUT	Neuter
NF	Non-future
NOM	Nominative
NOM1	Nominalized Verb Stem 1
NONFUT	Non-future
NP	Noun Phrase
O	Transitive object
OBJ	Object
OT	Optimality Theory
P	Preposition, Productivity
P^*	Global productivity
PART	Partitive
PAS	Predicate Argument Structure
PASS	Passive
PERF	Perfect(ive)
PERS	Person
PL	Plural
PP	Prepositional Phrase
PRES	Present
PRET	Preterite
PROG	Progressive
PTCP	Participle
Q	Qualifying
R	Relation, Relational
REL	Relative case
RHR	Right-hand Head Rule
S	Sentence, intransitive subject
s	strong
SC	Subject Concord

SG	Singular
SUBJ	Subject
SUPERL	Superlative
t	trace
TMA	Tense–Mood–Aspect
V	Verb, Vowel, number of types
VP	Verb Phrase
w	weak
WFR	Word Formation Rule
X, Y, x, y	variables
σ	syllable
ω	phonological word
<	derives from
>	results in, changes to
·	syllable boundary
-	morpheme boundary
´	primary stress, high tone, long vowel
`	secondary stress, low tone
$\Rightarrow$	is transformed into
$\rightarrow$	is realized as, is changed to
$\Leftrightarrow$	correlates with
>>	ranked higher than
[ ]	phonetic form
//	phonological form, underlying form

List of Figures

List of Tables

Preface to the Third Edition

In the third edition the suggestions for further reading have been updated, and a substantial number of references to important recent publications have been added.

The terminology used for inflection has been made more consistent, and some passages have been rewritten (with thanks to Rochelle Lieber for her constructive comments).

The Index of terms in previous editions has been expanded into a Subject index, as suggested by Barbara Schlücker.

Linguistic terms, printed in boldface when first used in the text, are explained in the Glossary.

Leiden, January 2012 G.B.

Preface to the Second Edition

I would like to thank the following colleagues and students for their remarks on the first edition of this book: Laurie Bauer, Jean-Marc Gachelin, Alexandra Galani, Istvan Kenesei, Elizabeth Koier, Francesca Masini, Franz Rainer, Iggy Roca, and Jeroen Wiedenhof. These remarks helped me in correcting typos and minor infelicities in the first edition. At the suggestion of Alexandra Galani, I have added a glossary of terms to this edition.

Leiden, July 2006 G. B.

Preface to the First Edition

Each textbook provides a specific perspective on the discipline that it aims to introduce. Therefore, writing this book has not only been a challenge for me because of the didactic demands that each textbook imposes on its writer. It also forced me to rethink my own ideas on morphology in confrontation with those of others, and to come up with a consistent picture of what morphology is about. This perspective is summarized by the title of this book, *The Grammar of Words*, which gives the linguistic entity of the word a pivotal role in understanding morphology.

It is with much pleasure that I would like to thank a number of colleagues for their constructive comments on an earlier draft of this book. Andrew Carstairs-McCarthy (University of Canterbury at Christchurch, New Zealand), Ingo Plag (University of Siegen), Sergio Scalise (University of Bologna), Caro Struijke (Vrije Universiteit Amsterdam), and Greg Stump (University of Kentucky at Lexington) read the whole manuscript, and gave extremely valuable advice. Maarten Mous and Marian Klamer (both University of Leiden) provided useful feedback for a number of chapters, and Mirjam Ernestus (Max Planck Institute for Psycholinguistics, Nijmegen) had a critical look at Chapter 10. Jenny Audring and Lourens de Vries (Vrije Universiteit Amsterdam), Maarten Kossman (University of Leiden), and Jaap van Marle (Open Universiteit Heerlen) also commented on a number of points. None of them should be held responsible for what I wrote in this book.

Another form of support for this project came from my colleagues at the University of Wisconsin at Madison, in particular Mark Louden, Monica Macaulay, and Joe Salmons. They made it possible for me to spend a very pleasant month in Madison, in which I could work on parts of this book.

It is my sincere hope that this textbook will prove to be useful for a new generation of students of language, and that they will enjoy reading and thinking about the many wonderful intricacies of human language.

Amsterdam G.B.

This page intentionally left blank

Part I

What is Linguistic Morphology?

This page intentionally left blank

1

Morphology: basic notions

1.1 Relations between words

When you use an English dictionary to look up the different meanings of the verb *walk*, you will not be surprised that there are no separate entries for *walk, walks*, and *walked*. You will also not feel disappointed if your dictionary does not contain a separate entry for *walking*. If you come across the sentence *My staff walked out yesterday*, and you want to find out what *walked out* means ("go on strike") you will not look for an entry *walked out*, but rather for an entry *walk out*. In many dictionaries, *walks, walked*, and *walking* are not even mentioned in the entry for *walk*. It is simply assumed that the language user does not need this information. The reason for the absence of this information is that these different English words are felt to be instantiations of the same word, for which *walk* is the citation form. So we have to make a distinction between the notion 'word' in an abstract sense (**lexeme**) and the notion 'word' in the sense of 'concrete word as used in a sentence'. The concrete words *walk, walks, walked*, and *walking* can be qualified as **word forms** of the lexeme WALK. Small capitals are used to denote lexemes when necessary to avoid confusion between these two notions 'word'. English dictionaries assume that the language user will be able to construct these different forms of the lexeme WALK by applying

the relevant rules. These rules for computing the different forms of lexemes are called rules of **inflection**.

This example shows that dictionaries presuppose knowledge of relations between words. It is the task of linguists to characterize the kind of knowledge on which the awareness of the relation between the word forms *walk, walks, walked*, and *walking* is based. Knowledge of a language includes knowledge of the systematicity in the relationship between the form and meaning of words. The words *walk, walks, walked*, and *walking* show a relationship in form and meaning of a systematic nature, since similar patterns occur for thousands of other verbs of English. The subdiscipline of linguistics that deals with such patterns is called morphology. The existence of such patterns also implies that words may have an internal constituent structure. For instance, *walking* can be divided into the constituents *walk* and *-ing*. Therefore, morphology deals with the internal constituent structure of words as well.

Dictionary makers assume that these forms of the lexeme WALK are formed according to rules, and therefore need not be specified individually in the dictionary. The same assumption plays a role in the case of nouns and adjectives. For English nouns, the plural form does not need to be specified in the dictionary if it is regular, and neither does the adverbial *-ly* form in the case of adjectives. For example, my English–Dutch dictionary (Martin and Tops 1984) does not mention the adverbs *correctly* and *economically* in addition to *correct* and *economical*. On the other hand, it does specify the adverb *hardly*. Why is that so? Is it due to inconsistency or sloppiness on behalf of the dictionary makers, or is there a principled reason behind this choice? There is indeed a reason: the meaning of *hardly* cannot be predicted from that of *hard* and *-ly*.

This kind of knowledge is also relevant when searching for information on the internet and in other digital data resources such as corpora of actual language use and electronic dictionaries. Suppose you want to collect information on tax. You might find it helpful if the search engine is programmed in such a way that it will not only recognize documents with the word *tax*, but also documents with the words *taxation, taxable*, and *taxability* as relevant. In fact, for many search engines this is not the case. The words *taxation* and *taxable* are both derived from the verb *to tax* which is related to the noun *tax*. The word *taxability* in its turn is derived from *taxable*. Hence, we may qualify this set of related words as a **word family**. On the other hand, when searching for information on tax

issues, you would not like your search engine to retrieve documents with the words *taxi, taxis, taxon*, or *taxonomy* that also begin with the letter sequence *tax*. This example shows that analysis of the systematicity in the relations between words is essential for the computational handling of language data. What we need for this purpose is a morphological parser, a computer program that decomposes words into relevant constituents: *tax-ation, tax-able*, and *tax-abil-ity*.

There is an intuitive difference between the members of the word family of TAX mentioned above and the set of word forms *walk, walks, walked, walking*. The different words related to the verb *to tax* are not felt as forms of the same word, but as different though related words that each have their own entry in the dictionary, that is, are different lexemes. We speak here of **lexeme formation** (or word-formation): TAXABILITY has been formed on the basis of TAXABLE through the addition of *-ity*, and TAXABLE in its turn has been formed on the basis of the verb TAX, just like TAXATION. The verb TAX itself has been formed by turning the noun TAX into a verb.

So far we have taken for granted that we can distinguish words from other linguistic units such as phrases, and we are no doubt influenced by the orthographical convention of using spaces to indicate word boundaries. Determining if a particular linguistic unit is a word is not always that easy, however, and certainly not for languages without a written tradition. Even for English we might not be certain. Why is *income tax* to be considered as a word rather than a phrase? After all, its constituents are separated by a space in its spelt form. The issue of word demarcation is taken up a number of times in this book.

Word-formation is traditionally divided into two kinds: **derivation** and **compounding**. Whereas in compounding the constituents of a word are themselves lexemes, this is not the case in derivation. For instance, *-ity* is not a lexeme, and hence TAXABILITY is a case of derivation. The word INCOME TAX, on the other hand, is a compound since both INCOME and TAX are lexemes. Changing the word class of a word, as happened in the creation of the verb *to tax* from the noun *tax*, is called **conversion**, and may be subsumed under derivation.

Another dimension of this kind of knowledge about words assumed by dictionary makers of English manifests itself in the fact that words that are quite common in English might not be covered by a dictionary. For instance, my English–Dutch dictionary does not mention *bottle factory*, although it does mention *bottle baby, bottle bank, bottleneck*, and a number

of other words beginning with *bottle*. Yet, I have no problem in understanding the title of the novel *The Bottle Factory Outing* written by Beryl Bainbridge. What the dictionary presupposes is that the user of English knows the words *bottle* and *factory*, and that the compound *bottle factory* refers to a particular kind of factory, not to a particular kind of bottle: it is the rightmost of the two word constituents that determines what kind of thing the compound denotes. This is a systematic fact of English. Therefore, one can understand the meaning of *bottle factory* without having ever come across that word before. That also applies to the even more complex word *bottle factory outing*. This example illustrates the creative aspect of morphological knowledge: it enables us to understand or coin new words. Morphological knowledge may thus lead to **rule-governed creativity** in the use of language. If we want to be understood, our new linguistic expressions must comply with the rules of the language. It is these rules that enable every language user to produce and understand linguistic expressions that she has never come across before.

The examples of morphological knowledge discussed so far come from English. The reason for this choice in an introductory chapter is a practical one: English is the language that all readers of this book are assumed to understand. English is not the obvious choice when one wants to discuss the nature of morphological systems in general, certainly not in the realm of inflection. After all, English has a relatively poor inflectional system, in which only a few grammatical distinctions are expressed. For instance, whereas English has only four different forms for regular verbs such as WALK, Romance languages such as French, Italian, and Spanish have tens of different forms for verbs. We should be aware of these considerable differences in morphological richness between languages. Therefore, it is important to look at a wide variety of languages in order to get a good idea of the morphological possibilities of natural language.

1.2 Paradigmatic and syntagmatic morphology

The term 'morphology' has been taken over from biology where it is used to denote the study of the forms of plants and animals. Its first recorded use is in writings by the German poet and writer Goethe in 1796. It was first used for linguistic purposes in 1859 by the German linguist August Schleicher (Salmon 2000), to refer to the study of the form of words.

In present-day linguistics, the term 'morphology' refers to the study of the internal structure of words, and of the systematic form–meaning correspondences between words. Consider the following sets of English words:

(1)	a.	buy	b.	buyer
		eat		eater
		paint		painter
		sell		seller
		send		sender

In these sets of words we observe a systematic form–meaning correspondence. The words in (1b) differ from the words in (1a) in that they have an additional part *-er*, and a corresponding meaning difference: each word in (1b) has the meaning "one who Vs", where V stands for the meaning of the corresponding verb in (1a). This is the basis for assigning a word such as *buyer* an internal morphological constituency: *buy-er*. The form differences between these two sets of words concern two properties: the words in (1b) have the additional sound sequence [ər] (or [ə] in standard British pronunciation) compared to the words in (1a), and they are nouns, whereas the words in (1a) are verbs. The form differences thus have a phonological and a syntactic dimension. The meaning difference is quite clear: the nouns in (1b) subsume the meaning of the corresponding verbs, and have some extra meaning due to the presence of *-er*. Since the nouns are formally and semantically more complex than the corresponding verbs, we will say that the nouns have been derived from the verbs. That is, there is a direction in the relationship between these two sets of words. The word *buyer* is a **complex word** since it can be decomposed into the constituents *buy* and *-er*. The word *buy*, on the other hand, is a **simplex word**, because it cannot be decomposed any further into smaller meaningful units, only into sound segments.

The notion 'systematic' in the definition of morphology given above is important. For instance, we might observe a form difference and a corresponding meaning difference between the English noun *ear* and the verb *hear*. However, this pattern is not systematic: there are no similar word pairs, and we cannot form new English verbs by adding *h-* to a noun. There is no possible verb *to heye* with the meaning "to see" derived from the noun *eye*. Therefore, such pairs of words are of no relevance to morphology. Similarly, we do not assign morphological constituency to German *fressen* "eating by animals" although it forms a pair with *essen* "to eat", since there

is no morphological constituent *fr-* that occurs in other word pairs as well. The words *fressen* and *essen* are in fact related historically (*fr-* derives from the early Germanic word *fra*), but *fressen* is no longer a complex word. So words can lose their status of complex word.

The existence of related words with a systematic form–meaning difference is crucial in assigning morphological structure to a word. The following Dutch words for different kinds of fish all end in *-ing*:

(**2**) bokking "bloater", haring "herring", paling "eel", wijting "whiting"

Yet, we do not consider this *-ing* a morphological constituent with the meaning "fish" because there are no corresponding Dutch words *bok, haar, paal*, and *wijt* with a meaning related to the corresponding words ending in *-ing* (these words do exist, but with a completely unrelated meaning).

The two sets of words given in (1) form **paradigms**. The term 'paradigm' is used here in a general sense to denote a set of linguistic elements with a common property. All words in (1a) are verbs, and thus form a paradigm. The same applies to the words in (1b) which are all nouns ending in *-er*. In our definition of morphology as given above we see two different perspectives. When we speak about morphology as the study of the systematic form–meaning correspondences between the words of a language, we take a paradigmatic perspective, since we take properties of classes of words as the starting point of morphological analysis. When morphology is defined as the study of the internal constituent structure of words, we take a syntagmatic perspective.

We distinguish these two different perspectives on language because language units exhibit **syntagmatic** and **paradigmatic relationships**. They have a syntagmatic relationship when they are combined into a larger linguistic unit. For instance, the words *the* and *book* have a syntagmatic relationship in the phrase *the book*. In contrast, the determiners *a* and *the* are paradigmatically related: they belong to the set of determiners of English, and can both occur at the beginning of a noun phrase, but never together: **the a book*. Hence, they belong to the paradigm of determiners of English.

A clear instantiation of a primarily syntagmatic approach to morphology is **morpheme-based morphology**. In this approach, focus is on the analysis of words into their constituent morphemes. That is, morphology is conceived of as the syntax of morphemes, as the set of principles for combining morphemes into words. **Morphemes**, the morphological building blocks of words, are defined as the minimal linguistic units with a lexical

or a grammatical meaning. For instance, the noun *buyer* consists of two morphemes, *buy* and *-er*. The verbal morpheme *buy* is called a **free** or **lexical morpheme**, because it can occur as a word by itself, whereas *-er* is an **affix** (hence a **bound morpheme** that cannot function as a word on its own). This is indicated by the hyphen preceding this morpheme: it requires another morpheme to appear before it in a word. Each of these morphemes is listed in the morpheme list of English: *eat* as a morpheme of the category Verb (V), and *-er* as an affixal morpheme of the category Noun (N) that is specified as occurring after verbs: [V —]. This specification of the affix *-er* assigns it to the subcategory of affixes that combine with verbs, and hence we call it a **subcategorization** property of this affix. The morphological structure of *eater* might be represented as follows:

(3) $[[\text{eat}]_{V}\ [\text{er}]_{\text{N-aff}}]_{N}$

This complex word can be created by the general mechanism of **concatenation**, the combination of elements into a linear sequence. This word is well formed because the requirement that *-er* occur after a verb is met. The fact that this combination of morphemes is a noun, and not a verb, follows from the generalization that English suffixes determine the category of the complex words that they create: since *-er* is an affixal noun, the whole word is a noun.

Thus, the language user is able to coin new **polymorphemic** words (words consisting of more than one morpheme) through the concatenation of morphemes, and of morphemes with words that are themselves polymorphemic. An example of the latter is the formation of the verb *tranquillize*, itself derived from *tranquil* through the addition of *-ize*. The formation of *tranquillizer* is not a matter of concatenating three morphemes. Instead, it is a two-steps operation. First, the bound morpheme *-ize* has been added to the simplex adjective *tranquil*, resulting in the verb *tranquillize*. Subsequently, the bound morpheme *-er* has been added to this verb. The morphological structure of this word is therefore a layered one, and can be represented in the form of a string with labelled bracketing, or as a tree (Figure 1.1). In short, morphology might be seen as morpheme syntax, as the set of principles that tell you how to combine free and bound morphemes into well-formed words.

This syntagmatic approach can be contrasted to a primarily paradigmatic approach to morphology. In the latter one, the creation of new complex words is seen first and foremost as the extension of a systematic

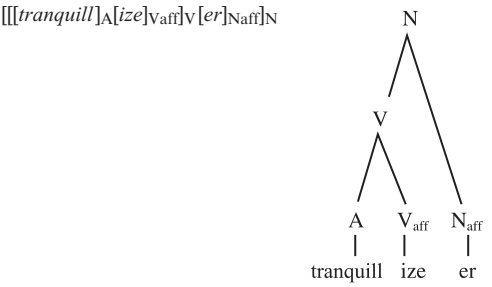

Fig. 1.1 The morphological structure of *tranquillizer*

pattern of form–meaning relationships in a set of established words to new cases, resulting in new words. Once we have discovered the abstract systematic pattern behind the words in (1), we will be able to extend this pattern to, for instance, the verb *swim*, resulting in the word *swimmer*:

(**4**) Pattern $[x]_V$: $[x\text{-}er]_N$ "one who Vs"; *swim : swimm-er*

(the variable *x* stands for the string of sound segments of the verb). In the gloss "one who Vs", the symbol V stands for the meaning of the verb. The gloss indicates that nouns ending in *-er* have a meaning that encompasses the meaning of the corresponding verb.

In this approach, it is not denied that the word *swimmer* consists of two constituent morphemes, but they are not the basic building blocks. Instead, words and relationships between words form the point of departure of the morphological analysis, and morphemes have a secondary status in that they figure as units of morphological analysis. Bound morphemes such as *-er* do not have lexical entries of their own, and only exist as part of complex words and of abstract morphological patterns such as (4).

From the point of view of language acquisition the paradigmatic perspective on complex words is the starting point of morphological analysis. When acquiring one's mother tongue, one has to discover the existence of morphological patterns on the basis of the individual words encountered in the input data. Only when language users have acquired a sufficient number of words of the relevant type, can they conclude to a systematic abstract pattern in sets of related words that might be used for the coinage of new words.

This paradigmatic pattern can receive a syntagmatic interpretation as well: the pattern can be interpreted as a morphological rule for the attachment of bound morphemes to words. That is, paradigmatic relationships

can be projected onto the syntagmatic axis of language structure. The pattern in (4) can thus be interpreted as the following morphological rule:

(5) $[x]_V \rightarrow [[x]_V\ er]_N$ "one who Vs"

This rule states that nouns with a particular meaning ("agent nouns") can be derived from verbal lexemes by adding the affix *-er* to the stem form of the verb. On the left side of the arrow, the requirements on the properties of the input words are specified, on the right side the formal and semantic properties of the output words. The arrow indicates the direction of the operation (input left, output right). The assumption of such affix-specific morphological rules means that bound morphemes do not exist as lexical items of their own, but only as part of morphological rules. Consequently, we get a slightly different representation of the morphological structure of the word *tranquillizer* given in Figure 1.2, without category labels for the affixes (compare Figure 1.1).

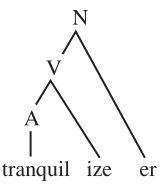

Fig. 1.2 The morphological structure of *tranquillizer* without affix labels

Instead of assuming a rule for this word-formation pattern, one might also express this regularity in the form of a **schema** for the coining of new nouns in *-er* of the following form that is formally equivalent to the morphological rule (5):

(6) $[[x]_V\ er]_N$ "one who Vs"

The morpheme-based approach and the lexeme-based approach may in fact lead to similar analyses of word structure. In both approaches, the polymorphemic noun *swimmer* will have the internal structure $[[swim]_V\ er]_N$. The (minor) difference is that in the word-based approach, the bound morpheme *-er* has no lexical category label of its own, since it is not a lexical entry. Yet, we should realize that rule (5), the rule interpretation of an extendable word pattern, has a paradigmatic flavour: it is not a rule about morpheme concatenation, but it specifies a formal and semantic operation (affix attachment and change of meaning) on lexemes. Similarly,

-schema (6) is an abstract scheme that specifies the common properties of a set of words, and can also be used as a 'recipe' to create new words.

It is essential that morphological rules can take **established** words as their inputs. If an established polymorphemic word has idiosyncratic properties, these properties will recur in words derived from it. For example, the complex noun *transformation* has a specific conventionalized meaning in generative syntax (the change of a syntactic structure). Hence, the same idiosyncratic meaning will recur in the adjective *transformational* derived from this noun *transformation*. Similarly, the adjective *edible* not only means that something can be eaten, but also that it can be eaten safely. This idiosyncratic meaning aspect of *edible* recurs in the derived noun *edibility*. Therefore, we must allow for established complex lexemes to function as the bases of word-formation. That is why morphological rules must be lexeme-based.

A particular challenge for the morpheme-based approach to morphology is the existence of morphological operations that do not consist of the concatenation of morphemes, so called **non-concatenative morphology**. The past tense forms of English irregular verbs, for instance, are not made through addition of a morpheme to a stem, but by replacement of vowels, as in *sing-sang*, and *grow-grew*. Another example (taken from Kutsch Lojenga 1994: 135) is that Ngiti, a Central-Sudanic language of Congo makes use of tones to distinguish morphologically related words. The plural form of a number of nouns is made by replacing the tones of the last two syllables (a sequence of a Mid tone and a Low tone) of the singular noun by a High tone. (The grave and acute accents indicate Low and High tones respectively; the absence of an accent indicates Mid tone.)

(7)

SINGULAR	PLURAL	
màlimò	malímó	"teacher(s)"
kamà	kámá	"chief(s)"
màlàyikà	màlàyíká	"angel(s)"

This process of forming plural nouns cannot be stated straightforwardly in a syntagmatic approach to morphology since there is no addition of a (tonal) morpheme. This pattern can be expressed straightforwardly in paradigmatic terms, as a systematic difference in form (tone pattern) correlating with the semantic distinction between singular and plural. Such a paradigmatic account of this regularity may look as in (8) where the schemas for singular and plural nouns of Ngiti are given:

(**8**) $[\ldots V^{M}.V^{L}.]_{Nsg}$, $[\ldots V^{H}.V^{H}.]_{Npl}$

The superscripts L, M, and H indicate the tone assigned to the vowel (V), and the Vs stand for the last two vowels of the words.

There are also cases of **paradigmatic word-formation**, in which a new word is formed by replacing one constituent of an existing complex word with another. For instance, the Dutch compound *boeman* "lit. boo-man" has a particular idiosyncratic meaning "ogre, bugbear". Its female counterpart *boevrouw* has obviously been coined by replacing the constituent *man* "man" with *vrouw* "woman" rather than by directly combining *boe* and *vrouw* into a compound, given the fact that the two compounds share this idiosyncratic meaning. Such a case of word-formation cannot be accounted for in a syntagmatic approach to morphology. It is based on specific words, and therefore a typical case of **analogy**:

(**9**) man : vrouw = boeman : *boevrouw* "female bugbear"

The paradigmatically oriented definition of morphology as the study of the systematic form–meaning correspondences between words expresses directly that morphology is lexeme-based. Lexemes form the point of departure of morphological processes. In lexeme formation (or word-formation) we create new lexemes on the basis of other lexemes, whereas in inflection, specific forms of lexemes are computed (instead of lexeme formation we will speak of word-formation when there is no risk of misunderstanding). The processes of word-formation and inflection together form the morphological part of a grammar.

Morphology deals with both the form and the meaning of linguistic expressions. Hence, one might qualify morphology as word grammar, that part of the grammar that accounts for the systematic form–meaning relations between words. In other words, it is a set of **correspondence rules** between forms and meanings of words. The notion 'word grammar' stands in opposition to 'sentence grammar', the grammar which describes the systematic relations between form and meaning at the sentence level.

1.3 The functions of morphology

The two basic functions of morphological operations are (i) the creation of new words (i.e. new lexemes), and (ii) spelling out the appropriate form of a lexeme in a particular syntactic context.

An example of the first function, lexeme formation, is given in section 1.1: the coining of the word *bottle factory* from the existing lexemes *bottle* and *factory*. Morphology thus provides means for extending the set of words of a language in a systematic way. The coinage of *bottle factory* is a case of compounding, in which two lexemes are combined into a new one. In the other type of word-formation, derivation, exemplified by the word *swimmer*, use is made of morphological operations on lexemes, whereas in compounding, two or more lexemes are combined into a new word.

Why do we need new words? One obvious reason is that language users need new expressions for new objects, or for new concepts. Once there is an entity or concept "factory for the production of bottles", it is quite convenient to be able to refer to such a concept with one word, *bottle factory* instead of using a circumscription. Thus, word-formation has a **labelling function**. Creating a word label for a new kind of entity, event, or property may have the additional pragmatic advantage that it draws attention to the new concept involved. For instance, the word *construction grammar* has been created to denote a particular school of linguistic thought in which the linguistic notion 'construction' plays a central role. By coining this label, a new linguistic school has been established, and thus its ideas will draw attention more easily. New verbs have been created to express new types of events or actions, such as the English verbs in *-ize: legal-ize* "to make legal", *tranquilll-ize* "to make tranquil", that express the causation of an event or property.

However, this is not the only function of word-formation. Another important function is that of syntactic **recategorization**: by using morphologically related words of different syntactic categories, we achieve stylistic variation and text cohesion, as the following examples (from Kastovsky 1986: 595) show:

(**10**) He made *fists* . . . He *defisted* to gesture.
If that's not *civil*, *civilize* it, and tell me.
[. . .] and whether our own conversation doesn't sound a little *potty*. It's the *pottiness*, you know, that is so awful.

A pragmatic reason for coining new words is found in the domain of evaluative morphology. In many languages **diminutive** forms of words are not used primarily for indicating the small size of the object denoted, but for giving a positive or negative evaluation. For instance, the Portuguese diminutive noun *avôzinho* (from *avô* "grandfather") means "dear grandfather" rather than "small grandfather", and in Dutch the diminutive noun

baantje "job" derived from *baan* "job" is used to refer to a job without prestige. A related phenomenon is that of the use of **attenuative** forms. The English morpheme *-ish* is often used to express the notion "sort of, not exactly": when we use *nine-ish* instead of *nine* as the time for an appointment, we mean that we do not expect people to be there at nine sharp. Thus, we can use morphology to express our subjective feelings towards something or to weaken or relativize a notion.

The function of inflection is primarily that of making forms of lexemes, including the correct forms of a lexeme appropriate for particular contexts. For instance, in an English clause, the verb has to agree with the subject with respect to number (singular or plural) and person (third or non-third), and this determines the choice between *walk* and *walks*: in a clause with present tense, *walks* has to be chosen if the subject is third person singular, and *walk* otherwise. In many languages, the form of a noun is determined by its syntactic context, and each noun has a number of **cases**. For instance, the Polish noun KOT "cat" has the case forms shown in (11). We call this structured set of word forms the *inflectional paradigm* of this lexeme (note that this is a more specific use of the notion 'paradigm' as introduced above in section 1.2). The term 'inflectional paradigm' may also be used to denote the abstract inflectional pattern, the set of labelled cells that these word forms occupy. As can be read off this paradigm of case forms, when the lexeme KOT occurs in direct object position and therefore has accusative case, the word form *kota* has to be used if the word has a singular meaning, and the form *koty* if it has a plural meaning. That is, one of the accusative forms has to be chosen for this syntactic position.

(**11**)

	SINGULAR	PLURAL	
NOMINATIVE	kot	kot-y	"cat, subject"
GENITIVE	kot-a	kot-ów	"of the cat"
DATIVE	kot-u	kot-om	"to the cat"
ACCUSATIVE	kot-a	kot-y	"cat, object"
INSTRUMENTAL	kot-em	kot-ami	"with the cat"
LOCATIVE	koci-e	kot-ach	"on the cat"
VOCATIVE	koci-e	kot-y	"o, cat"

Another function of morphology is that the relation between sentences in a text can be established by using morphological markers of **coreferentiality**. In Wambon, a language of New Guinea (examples from de Vries

1989: 62), verbal forms have Same Subject (SS) forms and Different Subject (DS) forms (1SG = first person singular, 3SG = third person singular, NF = non-future):

(12)	Nukhe	oye	khetak-mbel-o	topkeka-lepo
	I	pig	see-SS-COORD	flee-1SG.PAST
	"I saw a pig and I fled"			
(13)	Nukhe	oye	khetakha-lev-o	topkeka-tmbo
	I	pig	see-1SG.NF.DS-COORD	flee-3SG.PAST
	"I saw a pig and it fled"			

Both examples consist of two clauses, with the coordinating element *-o* linking these two clauses. This coordinative morpheme is attached to the verb of the first clause. The word *khetakmbelo* in (12) has the Same Subject form, which indicates that in both clauses we have the same subject "I". In (13), on the other hand, the word *khetakhalevo* is a Different Subject form, which indicates that the subject of the next clause is a different one. It is not "I", but the pig that fled. This kind of subject marking is called **switch reference**.

1.4 Morphology and the lexicon

The set of lexemes of a language comprises two subsets: simplex lexemes and complex lexemes. These lexemes are listed in the lexicon to the extent that they are established, conventionalized units. A complex lexeme like NINISH is a well-formed lexeme of English, but need not be listed in the lexicon since it is completely regular, and there is no conventionalization involved.

The lexicon specifies the properties of each word, its phonological form, its morphological and syntactic properties, and its meaning. The basic structure of lexical entries for the lexemes *swim* and *swimmer* may look as follows:

(14)	/swɪm/	/swɪmər/
	$[x]_V$	$[[x]_V \text{ er}]_N$
	$\text{SWIM}_{\text{ACTIVITY}}$	PERSON PERFORMING $\text{SWIM}_{\text{ACTIVITY}}$

The first line in these lexical entries specifies the phonological form of these lexemes: a sequence of sound segments between slashes. On the second line, categorial information, and internal morphological structure of a word are specified. On the third line, the meaning of the lexeme is specified, here

indicated by the use of small capital letters. The subscript ACTIVITY specifies the type of event expressed by this verb. A lexical entry thus expresses a correspondence between phonological, syntactic, and semantic pieces of information, just like morphological rules or templates, which do the same at a more abstract level, in a generalized fashion, with variables taking the place of the individual properties of lexemes.

Most complex words have been derived by one of the available word-formation processes of a language. Indeed, as we saw above, one of the main functions of morphology is to expand the set of available words. Once a complex word has been formed, it may get established as a word of the language. This means that it is used by more than one native speaker, and on different occasions, and that language users will recognize it as a word they have come across before. The set of established words of a language functions as the **lexical norm** or **lexical convention** of that language. For instance, in British English the machine that is used for drawing money from one's bank account is called a *cash dispenser*, and in American English it is called an *automatic teller machine (ATM)*. In fact, it would also have been possible to use the compound *money machine* for this device, but the established words function as a lexical norm, and hence they can block the creation of the compound *money machine*. That is, the lexicon as the set of established lexical units of a language may have a **blocking** effect on the creation of new words. It does not mean that *money machine* is an ill-formed word, only that its use might not be appropriate.

When a possible word has become an established word, we say that it has lexicalized. An important effect of **lexicalization** of complex words is that one of its constituent words may get lost, whereas the complex word survives. For instance, the Dutch verb *vergeet* "to forget" no longer has a simplex counterpart *geet*, unlike its English counterpart *forget* for which the corresponding word *get* does exist. We therefore consider *vergeet* a **formally complex word**. It still behaves as a complex verb since it selects a past participle without the prefix *ge-*, just like other prefixed verbs of Dutch. For example, the past participle of the prefixed verb *ver-wacht* "to expect" (derived from the verb *wacht* "to wait") is *verwacht*. Similarly, the past participle of the verb *vergeet* is *vergeten*, not **gevergeten*. This may be contrasted with the verb *verbaliseer* "to fine" in which the part *ver-* has no prefix status. The past participle of this verb is *ge-verbaliseer-d*, with the prefix *ge-* present.

The term 'lexicalization' is also used for a related phenomenon, namely that established words may have idiosyncratic, unpredictable, properties.

The meaning of *honeymoon*, for example, is not predictable on the basis of the meanings of its constituent lexemes *honey* and *moon*, and this requires this compound to be listed in the lexicon. Having idiosyncratic properties thus implies for a word that it has to be listed, but the inverse is not necessarily true: a complex word that is listed may have fully predictable properties, and may be listed only because it is an established word, that is, belongs to the lexical norm.

The notion 'lexicon' refers to the repository of all information concerning the established words and other established expressions of a language. It is an abstract linguistic entity, to be distinguished from the notion **dictionary**, which refers to practical sources of lexical information for the language user in some material (paper or electronic) form. A dictionary will never provide a full coverage of the lexicon due to practical limitations of size and requirements of user-friendliness, and because the lexicon is expanding and changing daily. The third related notion of relevance here is that of the **mental lexicon**, the mental representation of lexical knowledge in the brain of the individual language user. The mental lexicon of an individual is always smaller than the lexicon in the linguistic sense: nobody knows all the established words of a language. Moreover, the mental lexicon exhibits an asymmetry between production and perception: we understand probably about five times more words of our mother tongue than we actually use in language production.

In many languages, morphology is extremely important for the size of the lexicon. In all European languages, the number of established complex words is much higher than the number of simplex words. Consequently, the morphological rules of a language have two functions: they indicate how new lexemes and word forms can be made, and they function as **redundancy rules** with respect to the established complex words of a language. For instance, the lexical information that LOVER is a noun, and that the meaning of this word comprises that of the verb LOVE is redundant information. These properties are specified in rule (5). On the other hand, the information that this noun is an established word of English, with a particular idiosyncratic meaning "male sweetheart, suitor" is unpredictable, non-redundant lexical information.

Morphological patterns that can be systematically extended are called **productive**. The derivation of nouns ending in *-er* from verbs is productive in English, but the derivation of nouns in *-th* from adjectives is not: it is hard to expand the set of words of this type such as *depth, health, length, strength*, and *wealth*. Marchand (1969: 349) has observed some occasional coinings like *coolth* (after *warmth*), but notes that such word coinings are

often jocular, and hence do not represent a productive pattern. If we want to coin a new English noun on the basis of an adjective, we have to use *-ness* or *-ity* instead. In the case of unproductive patterns, the morphological rule involved functions as a redundancy rule only, and not as a rule for the creation of new words.

Lexical storage of complex morphological forms is also relevant in the realm of inflection. For example, Dutch has two plural endings for nouns, *-s* and *-en*. The second one is normally used for words consisting of one syllable. In the case of the monosyllabic noun *boon* "bean", the regular plural is *bon-en*, as expected. However, for *zoon* "son" both the irregular *zoon-s* and the regular *zon-en* can be used. Hence, the plural form *zoon-s* has to be specified in the lexicon.

The morphological system of a language is not its only source of complex words. There are at least three other sources: borrowing, phrases becoming words, and word creation.

As to **borrowing**, European languages have borrowed many words from Greek and Latin, often with French as the intermediary language. Consider the following list of Dutch verbs and their English glosses:

(**15**)	deduceer	"deduce"
	induceer	"induce"
	produceer	"produce"
	reduceer	"reduce"
	reproduceer	"reproduce"

A verb like *produceer* can be analysed into three parts: *pro-duc-eer*, that is, it is a polymorphemic word. The constituent *-eer* is a recurrent part of all these words, and so is *-duc-*. The sequences *de-*, *in-*, *pro-*, and *re-* are also recognizable elements in this set of verbs. Yet, we cannot say that these verbs have been created by a rule of Dutch or English morphology since there is no lexeme DUC from which these words could have been derived. Instead, a word such as *produceer* has been created by transforming the originally Latin verb *producere*, and by adapting its form by turning the ending *-ere* into *-eer*. The polymorphemic nature of such words remains recognizable in the borrowing languages. These borrowing patterns have led to a **pan-European lexicon**, a large stock of cognate complex words in the major languages of Europe.

A second non-morphological source of complex words is the **univerbation** ("becoming a word") of phrases. Phrases may lexicalize into words,

and thus lead to complex words. Examples from English are *jack-in-the-box, forget-me-not* (nouns), and *dyed-in-the-wool, down-at-heel, over-the-top* (adjectives). The following Dutch words all begin with *te-*, originally a preposition, the etymological cognate of English *to*:

(**16**) te-gelijker-tijd "lit. at same time, simultaneously"
te-rug "lit. to back, back"
te-vreden "lit. at peace, satisfied"
te-zamen "together"

In the first example, *tegelijkertijd*, the three constituents are clearly recognizable, and their meanings are relevant. The words *gelijk* "identical, same", and *tijd* "time" are current words of Dutch (the form of *gelijk* used here is *gelijker*, with an old inflectional ending *-er*). Therefore, *tegelijkertijd* is a complex, polymorphemic word. So the fact that a word is polymorphemic does not imply that it has been created by morphological rule. The second example, the word *terug*, is also interesting because it serves to illustrate a recurring problem of analysis for the linguist: when do we consider a word complex? Although *rug* "back" is a word of Dutch, it remains to be seen if we should consider *terug* a simplex or a complex word. In fact, many native speakers do not recognize the word *rug* in *terug* because of the more abstract meaning of *terug*, which no longer refers to a part of the human body.

Language users may also make new words by means of **word creation** (or **word manufacturing**). The following types can be distinguished:

(**17**) **blends:** combinations of the first part of one word with the second part of another: *brunch* < *breakfast* + *lunch*; *stagflation* < *stagnation* + *inflation*;
acronyms: combination of initial letters of a word sequence: *NATO* < *North Atlantic Treaty Organization*; French *OTAN* < *Organisation du Traité de l'Atlantique Nord*
alphabetisms: combination of the first letters of words, pronounced with the phonetic value of these letters in the alphabet: French *SVP* < *S'il vous plaît* "please"; Dutch *KLM* < *Koninklijke Luchtvaart Maatschappij* "royal airline company"; English *CD* "compact disc", *SMS* "Short Message Service";
clippings: one or more syllables of a word: *mike* < *microphone, demo* < *demonstration*, French *labo* < *laboratoire* "laboratory", German *Uni* < *Universität* "university".

In the case of compounds, only one of them may be shortened, as in German *U-Bahn* < *Untergrund-bahn* "metro", English *e-mail* "electronic mail", and *FAQ-list* "frequently asked questions list". In **ellipsis**, the first