

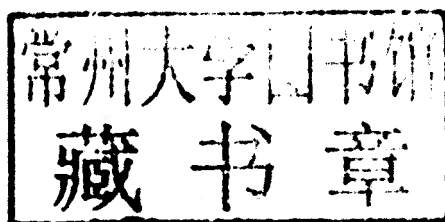
Construction Morphology

GEERT BOOIJ

OXFORD
UNIVERSITY PRESS

Construction Morphology

GEERT BOOIJ



OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

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 in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi

Kuala Lumpur Madrid Melbourne Mexico City Nairobi

New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece

Guatemala Hungary Italy Japan Poland Portugal Singapore

South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States

by Oxford University Press Inc., New York

© Geert Booij 2010

The moral rights of the author have been asserted

Database right Oxford University Press (maker)

First published 2010

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data available

Typeset by SPI Publisher Services, Pondicherry, India

Printed in Great Britain

on acid-free paper by

MPG Books Group, Bodmin and King's Lynn

ISBN 978-0-19-957192-5 (Pbk.)

ISBN 978-0-19-957191-8 (Hbk.)

1 3 5 7 9 10 8 6 4 2

Abbreviations and symbols

ˈ	main stress, high tone
ˌ	secondary stress, low tone
=	clitic boundary
-	affix boundary
↔	corresponds with
≈	is paradigmatically related to
·	syllable boundary
<	left edge of schema
>	right edge of schema
<	derives from
*	ungrammatical
<i>a</i>	variable for + or –
<i>ω</i>	phonological word
<i>σ</i>	syllable
Ø	zero
A	adjective
Acc	accusative
aci	accusativus-cum-infinitivo
Adj	adjective
Adv	adverb
Aff	affix
AP	adjectival phrase
CM	Construction Morphology
Comp	completive aspect
Cop	copula
D	determiner
Dat	dative
Def	definite
Dem	demonstrative
Det	determiner

Dig	digital
Dim	diminutive
DP	determiner phrase
Excl	exclusive
F	feminine, feature
Fem	feminine
Gen	genitive
Inf	infinitive
Inst	instrumental
Loc	locative
m	minimal number, non-plural
M	masculine, measure noun
Masc	masculine
MWE	multi-word expression
N	noun
Neg	negative
Neut	neuter
Nom	nominative, nominalization
NP	noun phrase
Num	numeral
Obj	object
Obl	oblique argument
Ord	ordinal
P	preposition, person, particle
pers	person
PHON	phonological representation
Pl	plural
Poss	possessive
PP	prepositional phrase
Pref	prefix
Pres	present tense
Progr	progressive
Ptcp	participle
R	relation
Refl	reflexive

RHR	right-hand head rule
SBJ	subject
SC	small clause
SCV	separable complex verb
SEM	semantic representation
Sg	singular
SOV	subject-object-verb
Subj	subject
Suff	suffix
SVO	subject-verb-object
SYN	syntactic representation
t	trace
TNS	tense
tr	transitive
V	verb
VN	verbal noun
VP	verb phrase
x,y	variable

Acknowledgements

This book is a synthesis of a number of ideas on and arguments for the use of the notion 'construction' in morphological analysis that I developed in recent years. In writing this book I have made use of some of my published articles on Construction Morphology. Some sections of Chapter 3 are taken from 'Compound construction: schemas or analogy? A Construction Morphology perspective', published in *Cross-disciplinary issues in compounding*, edited by Sergio Scalise and Irene Vogel, Amsterdam / Philadelphia: John Benjamins, 2010, 93–107. Chapter 4 is based on 'A constructional analysis of quasi-incorporation in Dutch', *Gengo Kenkyu* 135 (2009) 5–28, a journal published by The Linguistic Society of Japan. For Chapter 6, I have made use of my article 'Constructional idioms as products of language change: the *aan het* + INFINITIVE construction in Dutch', in Alexander Bergs and Gabriele Diebold (eds.) (2008), *Constructions and language change*, Berlin / New York: Mouton de Gruyter, 79–104. Chapter 7 is based on my article 'Phrasal names: a constructionist analysis' in *Word Structure* 2 (2009), 219–240, a journal published by Edinburgh University Press (www.eupjournals.com/jbctv). Chapter 8 has been written as a text for this book, but an adapted version has appeared as 'Constructions and lexical units: an analysis of Dutch numerals' in *Linguistische Berichte, Sonderheft* 19, 2010, 1–14, published by Helmut Buske Verlag, Hamburg. Finally, some of the data in Chapter 10 are from an article published in 2002, 'The balance between storage and computation in phonology' in Sieb Nooteboom, Fred Weerman, and Frank Weijnen (eds.), *Storage and computation in the language faculty*, Dordrecht: Kluwer, 115–138. The permission of each of the publishers to re-use this material is hereby gratefully acknowledged.

In preparing this book, I profited from the comments by anonymous reviewers on drafts of the publications mentioned above, and from those by my audiences on the occasions that I presented parts of this work.

Last but not least I would like to thank my colleagues who were kind enough to comment on a draft of the text of this book: Farrell Ackerman (San Diego), Kristel van Goethem (Leuven), Franz Rainer (Vienna), and Teresa Vallès (Barcelona). Their remarks led to various improvements, but of course the responsibility for this text is entirely my own.

Geert Booij
Leiden, January 2010

Contents

Abbreviations and symbols	viii
Figures and Tables	xi
Acknowledgements	xii
1. Morphology and construction grammar	1
1.1. Introduction	1
1.2. Word-based morphology	1
1.3. Constructions	11
1.4. Construction Morphology	16
1.5. Multi-word units	19
1.6. Inflectional patterns as constructions	22
1.7. Outlook	23
2. The lexicon as a network of relations	25
2.1. The hierarchical lexicon	25
2.1.1 Default inheritance	27
2.2. Paradigmatic relations	31
2.3. Constructional properties	36
2.4. Schema unification	41
2.4.1 Embedded productivity	47
2.5. A web of words	50
3. Schemas and subschemas in the lexicon	51
3.1. Schemas and subschemas	51
3.2. Semantic subgeneralizations require subschemas	55
3.3. Synchronic arguments for subschemas: headedness variation in compounding	66
3.4. Diachronic arguments for subschemas: <u>routes to affixhood</u>	71
3.5. Semantic arguments for constructional schemas	76
3.6. Subschemas for allomorphy patterns	84
3.7. Analogy or schema?	88
4. Quasi-Noun Incorporation	94
4.1. Morphological and syntactic constructions	94
4.2. Forms of noun incorporation	94
4.3. <u>Noun + verb</u> combinations in Dutch	97
4.4. Quasi-incorporation of bare plural nouns	108

4.5. Immobile verbs	112
4.6. Quasi-incorporation in Japanese	115
4.7. Conclusions	117
5. Separable complex verbs	118
5.1. Complex predicates	118
5.2. Dutch separable complex verbs	121
5.2.1. Lexical properties of particle verbs	125
5.3. SCVs as instantiations of constructional idioms	130
5.3.1. Preverb incorporation	134
5.3.2. Mismatches between form and meaning	137
5.4. Diachrony and grammaticalization	142
5.5. Conclusions	145
6. Progressive constructions	146
6.1. Periphrastic progressives	146
6.2. The <i>aan het</i> + INFINITIVE construction	150
6.3. Quasi-incorporation	159
6.4. Schema unification	163
6.5. The periphrastic role of the <i>aan het</i> + INF-construction	165
6.6. Grammaticalization	167
7. Phrasal names	169
7.1. Demarcating morphological and syntactic constructs	169
7.2. Naming and description	169
7.3. A + N combinations as names for concepts	175
7.3.1. Greek A + N combinations	179
7.3.2. Dutch A + N phrases with naming function	183
7.3.3. Syntactic lexical units and word formation	188
7.4. Theoretical implications: lexical phrasal constructions	190
8. Numerals as lexical constructions	193
8.1. Numerals: morphology or syntax?	193
8.2. Dutch numerals	193
8.3. Cardinal numerals	195
8.4. Ordinal numerals	205
8.5. Fraction numerals	206
8.6. Numerals and the architecture of the grammar	210
9. Construction-dependent morphology	211
9.1. Recycling morphology	211
9.2. The Definite -s Construction	216
9.3. The Partitive -s Construction	223

	CONTENTS	vii
9.4. Constructional idioms with -s		228
9.5. Collective constructions		231
9.6. Affixes as construction markers		235
10. Stem allomorphy and morphological relatedness		237
10.1. Underlying forms and lexical representations		237
10.2. The relation between morphology and phonology		239
10.3. Lexical phonological representations		242
10.3.1. The role of 'derived properties' of words		248
10.4. Stem allomorphy and relatedness of words		250
10.5. Stem allomorph selection		251
10.6. Stem allomorphy in the lexicon		253
11. Taking stock		255
References		260
Index		283

Figures and Tables

Figure 1.1. The tripartite parallel architecture	6
Figure 1.2. The lexical representation of <i>dog</i>	7
Figure 1.3. The lexical representation of <i>baker</i>	7
Figure 1.4. The schema for deverbal <i>-er</i>	8
Table 1.1. Examples of constructions varying in size and complexity	15
Table 1.2. The syntax–lexicon continuum	16

Morphology and construction grammar

1.1 Introduction

The title of this book, *Construction Morphology* (henceforth *CM*), promises a theory of linguistic morphology in which the notion ‘construction’ plays a central role. [The theory of *CM* aims at a better understanding of the relation between syntax, morphology, and the lexicon, and at providing a framework in which both the differences and the commonalities of word level constructs and phrase level constructs can be accounted for.]

In this chapter, I outline the main ingredients of this theory: a theory of word structure, a theory of the notion ‘construction’, and a theory of the lexicon. These are the topics of sections 1.2 and 1.3. In section 1.4 I discuss how the notion ‘construction’ can be made fruitful for morphological analysis and theorizing. A specific advantage of the notion ‘construction’ is that it can be used both at the level of word structure and that of syntactic structure without obliterating the differences between these two domains. This is shown in section 1.5 where phrasal units with word-like properties are introduced. Although this book focuses on word formation, inflectional phenomena also provide strong evidence for the correctness of a constructional approach, as briefly discussed in section 1.6. Section 1.7 provides a survey of the issues and phenomena that are discussed in the chapters that follow.

1.2 Word-based morphology

There are two basic approaches to the linguistic analysis of complex words. In the morpheme-based approach which was dominant in post-Bloomfieldian American linguistics, a complex word is seen as a concatenation of morphemes. In this approach, morphological analysis can be defined as the ‘syntax of morphemes’. For instance, the English word *walker* can be seen as a concatenation of the verbal morpheme *walk* and the nominalizing suffix *-er* that carries the meaning ‘agent’. This is the way in which English morphology

is often taught in textbooks, for example in Harley (2007). In a more radical form, the morpheme-based approach has even led to the claim that ‘morphologically complex words are the outcome of the manipulation of morphemes that takes place in syntax’ (Julien 2002: 297). Alternatively, we might take a word-based perspective in which words are the starting points of morphological analysis (Aronoff 2007). This is done by comparing sets of words like:

- (1) buy buyer
 eat eater
 shout shouter
 walk walker

We then conclude to a formal difference between the words in the left column and those in the right column. This difference correlates systematically with a meaning difference: the words on the right in (1) have an additional sequence *-er* compared to those on the left, and denote the agents of the actions expressed by the verbs on the left. Words like *buy* and *buyer* stand in a paradigmatic relationship, as opposed to the syntagmatic relationship that holds for words that are combined in a phrase or a sentence. This paradigmatic relationship between pairs of words like *buy* and *buyer* can be projected onto the word *buyer* in the form of word-internal morphological structure:

- (2) [[buy]_Ver]_N

In the mind of the speaker of English, the set of words listed in (1) may give rise to an abstract schema of the following (provisional) form:

- (3) [[x]_Ver]_N ‘one who Vs’

This schema expresses a generalization about the form and meaning of existing deverbal nouns in *-er* listed in the lexicon, and can also function as the starting point for coining new English nouns in *-er* from verbs. That is, new deverbal nouns in *-er* are not necessarily coined on analogy with a specific existing deverbal word in *-er*, but may be formed on the basis of this abstract schema. A new word is formed by replacing the variable *x* in the schema with a concrete verb. This is the operation of ‘unification’. For instance, the recently coined English verb *to skype* ‘to communicate by means of Skype’ can be unified with schema (3), resulting in the new noun *skyper*. As Tomasello (2000: 238) points out, language acquisition starts with storing mental representations of concrete cases of language use. Gradually, the language learner will make abstractions across sets of linguistic constructs with similar properties, thus acquiring the abstract system underlying these linguistic constructs.

The idea that word formation patterns can be seen as abstractions over sets of related words is rooted in a venerable tradition. For instance, the German linguist and Junggrammatiker Hermann Paul wrote in his famous *Prinzipien der Sprachgeschichte*, published in 1880, that [the language learner will start with learning individual words and word forms, but gradually (s)he will abstract away from the concrete words (s)he has learned, and coin new words and word forms according to abstract schemas]. This enables the language user to be creative in word formation and inflection (Paul 1880 [3rd edition 1898]: 102). This tradition is continued in the paradigmatic approach to word formation in the European tradition of word formation research (Schultink 1962; Van Marle 1985, 2000), in recent work in various varieties of non-transformational generative grammar such as Head-driven Phrase Structure Grammar (Riehemann 1998, 2001), and in the theoretical framework of Cognitive Linguistics (Croft and Cruse 2004; Langacker 1987, 1991; Taylor 2002).

Since such schemas depend on relationships between words, this morphological model has been called the network model (Bybee 1995), and the notion 'network' is indeed a proper term for conceptualizing the set of relationships between words in a lexicon (Bochner 1993). This approach may also be qualified as the 'abstractive' approach (Blevins 2006) because the coinage of new words depends on abstractions over sets of existing words and word forms in the lexicon of a language.

Schema (3) may be said to license the individual deverbal nouns in *-er* in the English lexicon. Complex words, once they have been coined, will be stored in the lexicon of a language (which generalizes over the lexical memories of the individual speakers of that language), if they have idiosyncratic properties and/or they have become conventionalized. A word is conventionalized if it has become the word to be chosen in a language community to denote a particular concept. For instance, the English compound *cash dispenser* is a word used to denote a machine from which one can take cash money. This machine can also be denoted by *cash machine* and *automatic teller machine* (ATM), but the word *money machine*, though well-formed and transparent as to its meaning, is not a conventional term for this device. Hence, words like *cash dispenser* must be stored in the lexicon.

This very short sketch of the analysis of a morphological pattern makes two assumptions. First, it assumes that there are specifically morphological generalizations or rules that cannot be reduced to either syntax or phonology. That is, this book takes the lexicalist position that the grammars of natural languages have a relatively autonomous morphological sub-grammar. Secondly, it assumes that complex words, i.e. the outputs of morphological operations, can be listed in the lexicon.

Morphological schemas have the following functions: they express predictable properties of existing complex words, they indicate how new ones can be coined (Jackendoff 1975), and they give structure to the lexicon since complex words do not form an unstructured list but are grouped into subsets. This conception of the grammar avoids the well-known rule versus list fallacy (Langacker 1987), the unwarranted assumption that linguistic constructs are either generated by rule or listed, and that being listed excludes a linguistic construct from being linked to a rule at the same time.

The relation between schema (3) and the individual words that conform to this schema is that of 'instantiation': each of the nouns in *-er* listed in (1) instantiates the schema in (3). Schema (3) provides a direct account of the fact that *-er* is a bound morpheme that does not occur as a word by itself, since this morpheme is not listed in the lexicon as an autonomous lexical item. Its existence is bound to its occurrence in schema (3). The same sequence of sounds /əɾ/ is used in other morphological schemas as well, for instance in the schema for the comparative form of English adjectives.

The use of constructional schemas like (3) looks similar to the use of word formation rules, as proposed in Aronoff (1976). The equivalent Aronovian rule is:

- (4) $[x]_V \rightarrow [[x]_V \text{er}]_N$ Semantics: 'one who Vs habitually, professionally'

The similarity between the two approaches is that they are both word-based (and hence affixes are not lexical items themselves), and both assume the coexistence of abstract patterns (rules/schemas) and complex words instantiating these rules/schemas listed in the lexicon. Yet, there are a number of advantages of schemas over rules that will be discussed in more detail in the next chapter. One difference that can already be mentioned here is that, whereas rules are always source-oriented (you take a base word, and perform some morphological operation on that base word), schemas can also be product- or output-oriented (Bybee 1995; Haspelmath 1989). For example, in Ngiti, a Central-Sudanic language of Zaire, the plural forms of nouns that are kinship terms or denote other inalienable possession are always characterized by a Mid-High tone pattern whereas the corresponding singular forms have a number of different tone patterns (Kutsch Lojenga 1994: 135) (Low tone is marked by ` , High tone by ' , and Mid tone is unmarked):

- | | | |
|-----|--------------------|---------------|
| (5) | <i>singular</i> | <i>plural</i> |
| | àba 'father' | abá |
| | abhu 'grandfather' | abhú |
| | adhà 'co-wife' | adhá |

Hence, the plural forms can only be characterized uniformly in terms of an output-oriented schema that specifies the Mid-High tone pattern of all these plural forms. The following schema is output-oriented, and expresses the relevant generalization:

- (6) [Mid High]_{N_i} 'plural N_i' (where N_i is inalienable)

The notion 'schema' is a very general notion from cognitive science. It is 'a data structure for representing the generic concepts stored in memory' (Rumelhart 1980: 34). That is, it can be used for making generalizations across all sorts of linguistic levels and types of (linguistic and non-linguistic) information. In Chapter 2 I discuss the properties of schemas in more detail in relation to the structure of the lexicon.

What is the implication of word-based morphology as outlined very briefly above for our conception of the architecture of the grammar? How does morphology fit into that architecture? My starting point is that each word is a linguistic sign, a pairing of form and meaning. The form of a word in its turn comprises two dimensions, its phonological form, and its morpho-syntactic properties. Hence, each word is a pairing of three types of information which will be labelled as PHON, SYN, and SEM respectively. Its meaning (SEM) may have both strictly semantic and pragmatic components (McConnell-Ginet 2008). Morphology affects all three dimensions of words. That is why we need a 'tripartite parallel architecture' of the grammar (as advocated by Culicover and Jackendoff 2005, 2006; Jackendoff 2002*a*, 2007 on the basis of primarily syntactic considerations). The essence of this model is that each level of representation is governed by rules and principles of its own, and that there are interface modules that specify the links between types of information on the different levels (Jackendoff 2002*a*: 125).

In Figure 1.1, Jackendoff uses the term 'rules' for regularities on a particular level of linguistic description, such as phonology or syntax. However, nothing hinges on this term, and one could use the term 'schema' here as well. For instance, for each language we need a phonological grammar that specifies how the sounds of a word are grouped into syllables and higher-level prosodic constituents such as the foot and the phonological word. The regularities in the phonological structure of words can be expressed by schemas for phonological structure, and the actual assignment of phonological structure to a word will then have the form of matching the sound sequence of that word with phonological schemas including those for prosodic structure. Hence, we might express the commonalities in the phonological properties of words as phonological schemas that generalize over the phonological properties of words. The notion 'schema' is a far more general notion than the notion

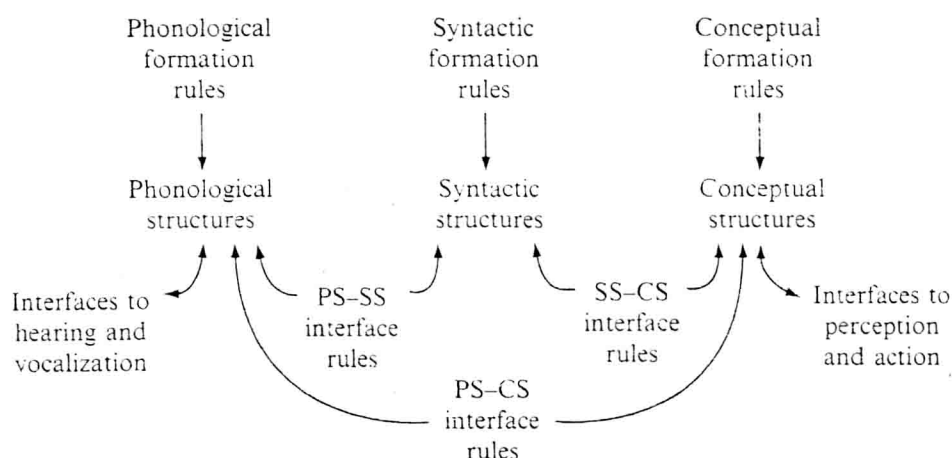


FIGURE 1.1. The tripartite parallel architecture

‘construction’ or ‘constructional schema’ which denotes a schematic pairing of form and meaning.

In sum, a word, like a sentence, is a complex piece of information. It links a particular sequence of sounds to a particular meaning, and it has formal properties such as a syntactic category label. The information contained in the English simplex word *dog*, for instance, can be represented in Figure 1.2, where the symbol \leftrightarrow stands for ‘correspondence’.

The first piece of information in Figure 1.2 concerns the phonological properties of this word: it is a phonological word (ω) that consists of one syllable (σ) that in its turn consists of a sequence of three sounds. This phonological word bears the same index as the syntactic information about this word (that it is a noun), and is also co-indexed with the semantic information that it expresses the predicate DOG. Co-indexation is used to specify the correspondence between the three kinds of information involved in knowing a word. We thus see that a word has a tripartite parallel structure.¹

¹ This is a traditional insight, formulated as follows by the linguist E. M. Uhlenbeck in his *dies* lecture for the University of Leiden in 1976: ‘woorden zijn eenheden waaraan drie dimensies zijn te onderkennen. Zij vertonen een hoorbare vorm – dit is hun fonische dimensie –, zij leveren in het gebruik een kennisbijdrage tot het geheel waarvan zij deel uitmaken – dit is hun semantische dimensie –, en tenslotte hebben zij een grammatische dimensie waaronder allereerst moet worden verstaan dat zij over systematische verbindingsmogelijkheden beschikken ten opzichte van andere woorden’ [words are units for which three dimensions can be distinguished. They exhibit an audible form – this is their phonic dimension –, they contribute knowledge to the expression as a whole – this is their semantic dimension –, and finally they have a grammatical dimension, which means first of all that they dispose of systematic possibilities of connection in relation to other words [my translation] (Uhlenbeck 1976)].