

word classes

OXFORD

LINGUISTICS

**word meaning**

*combinations*

structures **word sense**

**lexical knowledge**

# The Lexicon

An Introduction

Elisabetta Ježek

distribution

*ambiguity*

**context**

concept

OXFORD TEXTBOOKS IN LINGUISTICS

# The Lexicon

An Introduction

**Elisabetta Ježek**

**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

© Italian edition *Lessico. Classi di parole, strutture, combinazioni*  
first published by Il Mulino in 2005

© 2005 by Società editrice Il Mulino, Bologna  
English revised translation © Elisabetta Ježek 2016

The moral rights of the author have been asserted

This edition first published in 2016

Impression: 1

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, by licence 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 work in any other form  
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press  
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data  
Data available

Library of Congress Control Number: 2015940140

ISBN 978-0-19-960153-0 (hbk.)

ISBN 978-0-19-960154-7 (pbk.)

Printed in Great Britain by  
Clays Ltd, St Ives plc

Links to third party websites are provided by Oxford in good faith and  
for information only. Oxford disclaims any responsibility for the materials  
contained in any third party website referenced in this work.

OXFORD TEXTBOOKS IN LINGUISTICS

## The Lexicon

## OXFORD TEXTBOOKS IN LINGUISTICS

### PUBLISHED

- |   |   |
|---|---|
| The Grammar of Words                        | Compositional Semantics                           |
| An Introduction to Linguistic Morphology    | An Introduction to the Syntax/Semantics Interface |
| Third edition                               | by Pauline Jacobson                               |
| by Geert Booij                              |   |
| A Practical Introduction to Phonetics       | The History of Languages                          |
| Second edition                              | An Introduction                                   |
| by J. C. Catford                            | by Tore Janson                                    |
| Meaning in Use                              | The Lexicon                                       |
| An Introduction to Semantics and Pragmatics | An Introduction                                   |
| Third edition                               | by Elisabetta Ježek                               |
| by Alan Cruse                               | A Functional Discourse Grammar for English        |
| Natural Language Syntax                     | by Evelien Keizer                                 |
| by Peter W. Culicover                       | Diachronic Syntax                                 |
| Principles and Parameters                   | by Ian Roberts                                    |
| An Introduction to Syntactic Theory         | Cognitive Grammar                                 |
| by Peter W. Culicover                       | An Introduction                                   |
| A Semantic Approach to English Grammar      | by John R. Taylor                                 |
| by R. M. W. Dixon                           | Linguistic Categorization                         |
| Semantic Analysis                           | Third edition                                     |
| A Practical Introduction                    | by John R. Taylor                                 |
| by Cliff Goddard                            |   |
| Pragmatics                                  |   |
| Second edition                              |   |
| by Yan Huang                                |   |

### IN PREPARATION

- |   |                      |
|---|----------------------|
| Lexical Functional Grammar                      | Linguistic Typology  |
| by Mary Dalrymple, John Lowe, and Louise Mycock | Theory, Method, Data |
|   | by Jae Jung Song     |
| Translation                                     | Cognitive Grammar    |
| Theory and Practice                             | An Introduction      |
| by Kirsten Malmkjaer                            | 2nd edition          |
| Grammaticalization                              | by John R. Taylor    |
| by Heiko Narrog and Bernd Heine                 |                      |
| Speech Acts and Sentence Types in English       |                      |
| by Peter Siemund                                |                      |

## Preface to the English edition

The aim of this book is to provide the reader with the basic tools to approach the study of the lexicon. This is achieved by illustrating its structural organization and its behavior in actual use.

The lexicon of a language can be described as the set of its words. It will quickly become evident that to determine what constitutes a word in a language, to describe its meaning, how it is used, and how it can be combined with other words is no simple matter. On closer inspection, words prove to be a bundle of information of different sorts, and their meaning reveals multiple facets in context, inextricably linked with the syntax in which they appear.

We currently have a number of tools to scientifically describe the lexicon. Linguistic research, especially in the last 20 years, has devoted a lot of attention to this component of human language. This has occurred for several reasons, which we will not examine in detail—this book is intended to provide an introduction to the topic. Nevertheless, we should bear in mind that when we first learn a language we do so by initially using single words or unstructured sequences of words, and only later do we develop other skills, such as syntactic ones, often through the recognition and generalization of syntactic patterns associated with words. Moreover, if one engages in the study of human language from the point of view of the lexicon, one is forced to consider all aspects of language simultaneously (phonology, morphology, semantics, syntax and pragmatics), and to clarify how they interact. Words, in the end, are a direct mirror of how we conceptualize the world and our experience of it.

Since the study of the lexicon can be approached from several different perspectives, it is necessary to clarify the perspective taken in this book. This volume does not aim to present a theory of the lexicon from the point of view of its mental organization (i.e. the lexicon as it is stored and processed in our minds), nor does it aim to examine the way in which the lexical competence of speakers develops through time in language learning. More basically, its goal is to provide a description of the main properties of words and the organizational principles of the lexicon that can be derived by examining the use we

make of words on a daily basis; it introduces the categories that are useful to classify the various phenomena that can be observed by querying digitalized corpora and gathering speakers' judgments, and proposes viable representations for each of them, using the formalisms developed in the field of general and theoretical linguistics.

The book conforms to the tradition of structuralist and formal approaches to the study of language, and reexamines themes that have been a matter of debate since antiquity, such as the notion of word class, themes that found a systematization in the 1970s in the works of leading scholars such as J. Lyons and D. Dowty. In view of the lively debate surrounding current studies of the lexicon, the text provides an overview of central topics that are still highly controversial rather than a statement of a received doctrine.

The book is organized in two parts. The first part consists of the first three chapters. Chapter 1 presents the basics, introduces different lexicalization types, discusses what counts as a word in a language, and illustrates the types of words that exist in the languages of the world. Chapter 2 discusses the different types of information contained in words, and illustrates the interplay between lexical information and encyclopedic knowledge in the interpretation of words in their context of use. Chapter 3 is devoted to the meaning of words, and clarifies how words contribute to form the meaning of sentences, assuming a multi-dimensional model of lexical meaning.

The second part of the book consists of Chapters 4, 5, and 6. In this part, the focus shifts from the properties of individual words to an analysis of the structures that can be identified in the global lexicon. Chapter 4 discusses word classes. The focus is on nouns and verbs, for both of which several classes are proposed, identified on the basis of both syntactic and semantic criteria (valency structure, *Aktionsart*, type of entity, etc.). The last part of the chapter focuses on typological variation, and presents the main systems of word classes that are attested in the world's languages, based on Hengeveld's model. Chapter 5 discusses paradigmatic relations between words based on their meaning (synonymy, hyponymy, etc.), whereas Chapter 6 introduces the main constraints operating on the combination of words, and describes the main types of possible combinations.

I would like to thank the following colleagues who read the original manuscript or parts of it for their valuable comments: Giorgio Arcodia, Nicola Grandi, Claudio Iacobini, Alessandro Lenci, Michele Prandi, Paolo Ramat, Raffaele Simone. Raffaele Simone deserves a special thanks for his

advice on the content and structure of the book. I am also grateful to James Pustejovsky, whose work on lexical and compositional semantics has greatly inspired my interest in the study of the lexicon, and to Patrick Hanks for introducing me to the use of corpora in lexical analysis. A special thanks also goes to the students of the Syntax and Semantics class of the Master Program of Theoretical and Applied Linguistics I run at Pavia University, whose comments helped me refine the presentation of several concepts for pedagogical purposes. Finally, I acknowledge Jaap Hillmann for his constant support during the preparation of the English edition and Peter Lofthouse for his most valuable feedback on the English language. Family, friends, and acquaintances deserve my warm thanks for supporting the effort required to complete this project. Thanks also to four anonymous reviewers for their comments and to the editorial staff of OUP for their help in production. All errors are, of course, my own.

I would like to dedicate this book to the memory of my father, Giorgio Ježek.



# Typographic conventions

## Asterisks

For expressions that are semantically incongruent and/or grammatically incorrect, e.g. \*a well meal, \*an evening black dress.

## Question marks

For expressions whose semantic or grammatical acceptability is uncertain, e.g. ? The man looked *quickly* up.

## Small capitals

For concepts and ontological categories, e.g. PERSON, THING, PLACE.

For semantic primitives, e.g. DO, CAUSE, BECOME.

For lexical relations, e.g. HAS\_AS\_PART, HAS\_AS\_PREREQUISITE.

## Capitals

For semantic primes in Wierzbicka's model, e.g. PEOPLE, UNDER, NOW, VERY.

## Round Brackets

For semantic and aspectual features, e.g. (male), (abstract), (homogeneity).

## Angle brackets

For selectional restrictions, e.g. <animate>, <eatable>.

## Bold Type

For technical terms when first defined, e.g. **frame**, **qualia**.

## Italics

For technical terms introduced but not defined, e.g. we will examine the notion of *lexicalization* in section 2.

For lexemes and expressions when used in the text, e.g. *simple*, *be late*, *wine glass*.

For emphasis on specific lexemes in examples of use, e.g. "a *do-it-yourself* culture."

## Single quotation marks

For meanings, e.g. Chinese *bing* may mean 'get sick' (verb) but also 'illness' (noun).

Double quotation marks

For examples of use, e.g. "a comfortable chair."

For examples from corpora, e.g. "The species' classification evolved during time from static to dynamic."

For emphasis on terms, e.g. the "legitimate" parts of an object.

For quotations from other authors, e.g. "the categories and types that we isolate from the world of phenomena we do not find there because they *stare every observer in the face*" (Whorf 1956).

Square brackets

For phonetic representation, e.g. [i:] for *bee*.

## Abbreviations

adj.	adjective
adv.	adverb
art.	article
CLF	classifier
det.	determiner
Engl.	English
f.	feminine
Fr.	French
Gen.	genitive
Germ.	German
Gr.	Greek
ingr.	ingressive
It.	Italian
Lat.	Latin
m.	masculine
n.	neuter
Nor.	Norwegian
obj.	object
perf.	perfect
plur.	plural
prep.	preposition
sing.	singular
subj.	subject

# Contents

Preface to the English edition viii

Typographic conventions xi

Abbreviations xiii

## 1. Basic notions 1

1.1. Lexicon and dictionary 1

1.2. Lexicon and semantics 5

1.2.1. Concepts and lexicalization 5

1.2.2. Lexicalization patterns 7

1.3. Lexical meaning and grammatical meaning 13

1.3.1. Content words and function words 14

1.3.2. Types of grammatical meanings 16

1.4. The notion *word* 21

1.4.1. Counting the words of a language 22

1.4.2. Word types 29

1.4.3. Word typology 35

Further reading 39

## 2. Lexical information 41

2.1. Types of lexical information 41

2.2. Lexical information and encyclopedic knowledge 50

Further reading 53

## 3. The meaning of words 54

3.1. What, exactly, do words mean? 54

3.2. Ambiguity and polysemy of words 56

3.3. Theories on the nature of meaning 62

3.4. Building meaning syntagmatically 75

3.5. Systems of word meaning representation 83

Further reading 97

## 4. The global structure of the lexicon 98

4.1. Word classes 98

4.2. Morphological, syntactic, and semantic classes 101

- 4.3. Relationships between word classes and meaning 102
- 4.4. Word subclasses 107
  - 4.4.1. Verb classes 107
    - 4.4.1.1. Transitivity and intransitivity 108
    - 4.4.1.2. Valency structure 111
    - 4.4.1.3. *Aktionsart* 121
    - 4.4.1.4. The root 127
  - 4.4.2. Noun classes 129
    - 4.4.2.1. Entity nouns 130
    - 4.4.2.2. Event nouns 136
- 4.5. Word class systems across languages 151
- Further reading 157
- 5. Paradigmatic structures in the lexicon 159**
  - 5.1. What counts as a paradigmatic relation? 159
  - 5.2. Kinds of paradigmatic relations among words 163
  - 5.3. Relations of inclusion: hyperonymy/hyponymy; meronymy/holonymy 164
  - 5.4. Relations of identity: synonymy, near-synonymy 172
  - 5.5. Relations of opposition: antonymy, complementarity, converse terms 175
  - 5.6. Other relations: cause, purpose, temporal entailment 180
  - 5.7. Lexical configurations 182
  - Further reading 184
- 6. Syntagmatic structures in the lexicon 185**
  - 6.1. What is a syntagmatic relation? 185
  - 6.2. Syntagmatic relations: role, attribution, and manner 186
  - 6.3. Impossible word combinations: why? 188
    - 6.3.1. Selectional restrictions and lexical solidarities 189
    - 6.3.2. Types of constraints on word combination 190
      - 6.3.2.1. *Context fitting* 194
      - 6.3.2.2. *Redundancy* 194
  - 6.4. Types of word combinations 195
    - 6.4.1. Free combinations 197
    - 6.4.2. Restricted combinations 198
    - 6.4.3. Collocations 199
      - 6.4.3.1. *Salience vs. frequency* 203
    - 6.4.4. Light verb constructions 204

6.4.5. Other constructions	207
6.4.6. Idiomatic expressions	208
6.5. Impact of combinatory phenomena on the lexicon	210
Further reading	215

Concluding observations	216
-------------------------	-----

References	217
------------	-----

Index	229
-------	-----

# 1

## Basic notions

1.1 Lexicon and dictionary	1	1.4 The notion <i>word</i>	21
1.2 Lexicon and semantics	5	Further reading	39
1.3 Lexical meaning and grammatical meaning	13		

To address the study of the lexicon a list of basic notions are needed; these are provided in the first part of this chapter. In the second part, we examine the notion of *lexicalization*, and discuss the many ways in which information content may be associated with a word. Finally, we address the question of what counts as a *word* in a language and illustrate the different types of words that can be found in languages throughout the world.

### 1.1. Lexicon and dictionary

The **lexicon** is the set of words of a language, while the **dictionary** is the work of reference that describes that word set. Lexicon and dictionary do not correspond. A dictionary is a concrete object, typically a book, in either printed or electronic format, whereas the lexicon is an abstract object, that is, a set of words with associated information, stored in our mind and described in the dictionary. The relationship between these two entities is approximately the same as the relationship that exists between the grammar of a language, understood as the set of its syntactic and morphological rules, and the book that lists these rules and illustrates how they apply (a grammar of English, of Italian, of Hindi, and so on).

The structure of the lexicon does not correspond to the structure of the dictionary. In dictionaries, the information is organized according to practical considerations; for example, words are listed in alphabetical order for ease of consultation. The lexicon, instead, is not organized alphabetically but rather in word groupings whose members share similarities from the point of view of their form and/or their meaning. Typical lexical structures are, for example: morphological word families, such as *book*, *booking*, *booklet*, *bookstore*, based on the presence of the word *book*; semantic networks such as *buy*, *acquire*, *purchase*, *sell*, *negotiate*, *pay*, *own*, based on meaning associations; and groups of words with similar syntactic behavior, for example *nouns*, *verbs*, or *adjectives*. The dictionaries that mimic more closely the structure of the lexicon are the so-called “semantic” dictionaries, which organize words on the basis of the proximity of their meaning and put for example *nail* near *hammer*, *concert* near *music*, *paint* near *paint brush*, *table* near *chair*, *kitchen* near *eat*, *park* near *garage*, and so forth. Since Roget’s work (*Thesaurus of English Words and Phrases*, 1852), these “semantic” dictionaries are called **thesauri**, using a word borrowed from Greek via Latin. Standard thesauri list only words which are similar with respect to a given heading (for example *ship*, *vessel*, *boat*, *craft*) while extended versions also include word groupings which are based on different kinds of relations, such as those holding between the words mentioned above (a *paint brush* is used for *painting*, what you typically do in the *kitchen* is *eat*, and so forth).

Finally, the term **vocabulary** may be used to refer to both the body of words in use in a particular language (hence, its lexicon), and the reference work that collects and describes this heritage (therefore, its dictionary).

We have said that a dictionary is the description of a lexicon. In fact, it is more appropriate to say that it is an attempt to describe the lexicon, just as a grammar book is an attempt to describe the syntactic and morphological rules of a language, which we often experience as lacking or imprecise in many respects. The very fact that multiple dictionaries or grammars are available for the same language shows that both the lexicon and the rules of grammar can be described according to various principles and from different perspectives. The dictionary never constitutes an exhaustive source of all the words, meanings and attested or possible usages of words in a language.

In many cases, the incompleteness of dictionaries is a deliberate choice. For example, a dictionary of contemporary usage might intentionally exclude rare, literary, archaic, or specialized words. In other



cases, however, incompleteness is the unintended consequence of the fact that the number of words in a language is hard to determine (section 1.4.1) and the properties of the individual words are not so easily identifiable (Chapter 2). Finally, we must consider that the information to be found in dictionaries, despite being incomplete, is usually larger than the average lexical competence of an individual speaker. Indeed, a native speaker never knows all the words, the meanings, and the uses documented in a dictionary, much less specific information such as, for example, the etymology of a word and its first documentation in texts. Conversely, the lexical competence of a native speaker is not merely a sub-set of the information contained in dictionaries. Missing from dictionaries are complex words with highly predictable meaning, such as Dutch *winkeltje* ‘small shop,’ formed on the basis of *winkel* ‘shop’ through the addition of *-tje* ‘small.’ Such forms are often not included in dictionaries based on the idea that they do not reside permanently in the speaker’s lexicon, especially if they are not frequent in use, but are constructed “online” by speakers by directly combining items in the context of use.

The disciplines that study the lexicon and the dictionary also have different names and purposes. **Lexicology** investigates the lexicon of a language with the goal of identifying the inherent properties of words and illustrating how, by virtue of their meaning, words relate to one another and may be successfully combined into larger expressions that are both meaningful and grammatical. In its most recent developments, lexicology aims at emphasizing that both the information associated with words (Chapter 2) and the lexicon in its entirety (Chapters 4–6) are highly structured. This view contrasts with the idea that the lexicon, unlike grammar, is a loosely structured collection of information—an idea that has circulated in the linguistic community for a long time, according to which the lexicon is “a list of basic irregularities.” Given the composite nature of lexical knowledge (which can be said to include phonologic, morphological, semantic, and syntactic knowledge, cf. Chapter 2), lexicology is a discipline that exploits methodologies and formalisms developed in related fields of study. These include **lexical semantics**, concerned with defining the meaning of words, explaining its flexibility in context, and accounting for how it contributes to the meaning of sentences, or syntax, concerned with how words may or may not be combined, their relations of dependence, etc. The final goal of lexicology is to develop a **theory of the lexicon**, that is, a hypothesis about