LEXICAL FUNCTIONS IN LEXICOGRAPHY AND NATURAL LANGUAGE PROCESSING

Edited by

LEO WANNER
University of Waterloo

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

LEXICAL FUNCTIONS IN LEXICOGRAPHY AND NATURAL LANGUAGE PROCESSING

Edited by

LEO WANNER
University of Waterloo

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences — Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Lexical functions in lexicography and natural language processing / edited by Leo Wanner.
p. cm. -- (Studies in language companion series, ISSN 0165-7763 ; v. 31)
Most of the articles in this book grew out of the papers presented at the International Workshop on the Meaning-Text Theory, held at the Institut für Integrierte Publikations- und Informationssysteme in 1992.
Includes bibliographical references and indexes.
1. Lexicography. 2. Lexicology. 3. Computational linguistics. 4. Functionalism (Linguistics) I. Wanner, Leo. II. International Workshop on the Meaning-Text Theory (1992 : Institut für Integrierte Publikations- und Informationssysteme) III. Series: Studies in language companion series ; v. 31.
P327.L419 1996
413';028--dc20 95-53207
ISBN 90 272 3034 X (Eur.) / 1-55619-383-1 (US) (alk. paper) CIP

John Benjamins Publishing Co. • P.O.Box 75577 • 1070 AN Amsterdam • The Netherlands
John Benjamins North America • P.O.Box 27519 • Philadelphia, PA 19118 • USA

Preface

The goal of this book is to summarize the most important work carried out on and with Lexical Functions in the field of lexicography and computational linguistics.

Most of the articles in this book grew out of the papers presented at the *International Workshop on the Meaning-Text Theory* held at the *Institut für Integrierte Publikations- und Informationssysteme*, Darmstadt, Germany in July 1992. The workshop was funded by the German Research Association (*Deutsche Forschungsgemeinschaft*) and organized by Karin Haenelt and myself.

Each article has been reviewed by two other contributors to the volume. Additionally, Richard Kittredge, Tanya Korelsky, and Elke Teich were so kind as to take on the thankless job of reviewing. Bruce Jakeway, Lidija Iordanskaja, Igor Mel'čuk, and Alain Polguère did the proofreading. I would like to express my heartfelt gratitude to all who helped to ensure a high quality book on, as I think, a very interesting and important topic.

Leo Wanner

Authors

Margarita Alonso Ramos
Linguistica Xeral
Universidade da Coruña
Campus da Zapateira s/n
15071 Coruña
Spain

Martha W. Evens
Department of Computer Science
Illinois Institute of Technology
10 West 31st Street
Chicago, IL 60616
U.S.A.

Joseph E. Grimes
Summer Institute of Linguistics
and Cornell University
84-664 Ala Mahiku Drive, 191-B
Waianae, HI 96792
U.S.A.

Ulrich Heid
IMS, Computerlinguistik
Universität Stuttgart
Azenbergstr. 12
D-70174 Stuttgart
Germany

Lidija N. Iordanskaja
CoGenTex
5913 Rue Dolbeau
Montréal, QC
Canada H3S 2G1

Myunghee Kim
CoGenTex
5913 Rue Dolbeau
Montréal, QC
Canada H3S 2G1

Woongjae Lee
Department of Computer Science
Seoul Woman's University
126 Kognung-2 Dong, Nowon-Ku
Seoul
Korea 139-774

Igor A. Mel'čuk
Département de Linguistique
Université de Montréal
C.P. 6128 "Centre Ville"
Montréal, QC
Canada H3C 3J7

Alexander Nakhimovsky
Computer Science Department
Colgate University
Hamilton NY, 13346
U.S.A.

Natal'ja Percova
Scientific Research Center
Moscow State University
Vorobjevy Gory
119899 Moscow
Russia

Alain Polguère
Département de Linguistique
Université de Montréal
C.P. 6128 "Centre Ville"
Montréal, QC
Canada H3C 3J7

Agnès Tutin
Laboratoire SILEX, URA 382
Université de Lille III
59653 Villeneuve d'Ascq Cedex
France

Tilmann Reuther
Institut für Slawistik
Universität Klagenfurt
Universitätsstr. 65–67
A-9020 Klagenfurt
Austria

Leo Wanner
Computer Science Department
University of Waterloo
Waterloo, ON
Canada N2L 3G1

Abbreviations and Notations

A	: adjective
ABL	: Ablative (case)
ACC	: Accusative (case)
ACT	: active (voice)
ADV	: adverb
-A	: actant
ART	: an article or a grammatically equivalent determiner
AP	: adjectival phrase
ATTR	: deep-syntactic relation "Attribute"
BBI	: The BBI Combinatory Dictionary of English (Benson *et al.*, 1986)
C, C_0	: keyword of a given Lexical Function
CED	: Collins English Dictionary
CONJ	: conjunction
CQP	: Corpus Query Processor
CV	: consonant + vowel
CVV	: consonant + long vowel
DAT	: Dative (case)
DER	: (Deep-Syntactic) Derivative
DET	: determiner
DSynt	: Deep-Syntactic
DSyntA	: Deep-Syntactic Actant
DSyntR	: Deep-Syntactic Representation
DSyntRel	: Deep-Syntactic Relation
DSyntS	: Deep-Syntactic Structure
ECD	: Explanatory Combinatorial Dictionary
f, **g**, **h**	: a given LF
$G_{1,\ldots,8}$	: a class of grammatical collocations in (Benson *et al.*, 1986)
GER	: gerund
GP	: Government Pattern [$\approx$ subcategorization frame]
GEN	: Genitive (case)
GF	: grammatical function

INF	: infinitive
INT	: interjection
kwic	: keyword in context (concordancing program)
$\mathcal{L}$	: a given language
L	: a given lexical unit
L_{gen}	: the generic lexical unit of a given semantic field
$L_{1,\ldots,7}$	: a class of lexical collocations in (Benson *et al.*, 1986)
LF	: Lexical Function
LU	: lexical unit
MAS	: Malyj Akademičeskij Slovar′
MTM	: Meaning-Text Model
MTT	: Meaning-Text Theory
N	: noun
NLP	: Natural Language Processing
NOM	: Nominative (case)
NP	: noun phrase
NUM	: numerical
OALD	: Oxford Advanced Learner's Dictionary
OŽ	: Ožegov and Švedova's dictionary (Ožegov & Švedova, 1992)
PASS	: passive (voice)
PERF	: perfective reading of a verb
poss	: possessive (pronoun)
PRON	: pronoun
PP	: prepositional phrase
PREP	: preposition
$PREP_{dir}$	: directional preposition
PROP	: proposition
RedSemR	: Reduced Semantic Representation
S	: substantive (=DSynt-noun)
Sem	: Semantic
SemA	: Semantic Actant
SemR	: Semantic Representation
SemS	: Semantic Structure
SFL	: Systemic Functional Linguistics
sg	: singular
SPhonR	: Surface-Phonetic Representation
SSynt	: Surface-Syntactic
SSyntR	: Surface-Syntactic Representation
SSyntRel	: Surface-Syntactic Relation
SSyntS	: Surface-Syntactic Structure

SV	: support verb
SVC	: support verb construction
TKS	: Tolkovo-Kombinatornyj Slovar' 'Explanatory Combinatorial Dictionary (of Russian)'
UŠ	: Ušakov's dictionary (Ušakov, 1935)
V	: verb
VB	: verb (basic form)
VP	: verb phrase
X, Y, Z, W	: semantic actant denotation
X | Y	: Y is the condition for X
'X'	: the meaning of X
$\ulcorner X_1+X_2+ \ldots +X_n \urcorner$	: idiom or quasi-idiom composed of LUs $X_1+X_2+ \ldots +X_n$
I, II, III, IV	: deep-syntactic actant denotation
1, 2, 3, 4	: surface-syntactic actant denotation
[1],[2],[3],...	: endnote numbering
$\sim$	: placeholder for the head LU (in a lexical entry), alternative
$\approx$	: equivalence
$\neg$V	: negation of an inherited value V of a given LF
$\uparrow$**f**	: add the value of the LF **f** to the inherited set of the values of **f**

Contents

Abbreviations and Notations xvii

Introduction 1

Leo Wanner

1 On LF Relations . . . 3

1.1 Paradigmatic LF Relations . . . 3

1.1.1 LF Relations and Sense Relations . . . 4

1.1.2 LF Relations and Syntactic Paradigmatic Relations . . . 10

1.1.3 LF Relations and Argument Roles . . . 11

1.2 Syntagmatic LF Relations . . . 12

1.2.1 Typology of Syntagmatic Relations . . . 13

1.2.2 Introduction to Collocational Relations . . . 15

1.2.3 Collocational LF Relations *vs.* Collocation Classes . . . 18

2 Lexical Functions in the Overall Framework of MTT . . . 23

3 Outline of the Volume . . . 25

Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon 37

Igor Mel'čuk

1 Preliminary Remarks . . . 37

2 On the Concept of Lexical Function . . . 39

3 Simple Standard Lexical Functions . . . 45

3.1 Classification of Simple Standard LFs . . . 46

3.2 List of Simple Standard LFs . . . 46

3.2.1 Paradigmatic LFs (nos. 1–25) . . . 47

3.2.2 Syntagmatic LFs (nos. 28–64) . . . 56

4 Special Phenomena Related to Lexical Functions 72
4.1 Complex LFs 73
4.2 Configurations of LFs 74
4.3 Fused Elements of Values of LFs 74
5 Presentation of the Values of LFs in Lexical Entries 74
5.1 Elements of the Value f(L) as Subentries in L's Entry 75
5.2 Generalizing over the Values of LFs 76
5.2.1 Shared Values of LFs and Cross-References to a Convenient LU 77
5.2.2 Shared Values of LFs and the Generic LU of the Semantic Field 77
5.2.3 Shared Values of LFs as Universal Elements of LF Values 79
5.2.4 Shared Values of LFs as Separate Lexical Entries 79
6 Linguistic Nature of Lexical Functions 79
6.1 Semantic Aspect of LFs 80
6.1.1 LFs Are Not Semantic Units 80
6.1.2 LFs and Their Relation to the Definitions . 82
6.1.3 Semantic Subtypes of LFs 83
6.1.4 Different Semantic "Facets" of a Keyword . 85
6.1.5 LFs as Deep-Lexical Units 85
6.2 Phraseological Aspect of LFs 86
6.2.1 Graduality of Restrictedness of LF Expressions 86
6.2.2 LF Expressions and Collocations 88
6.3 LFs in Linguistic Representations 88
6.4 Universality of LFs 89
6.5 New Simple Standard LFs? 90
7 Lexical Functions in Computer Applications 91
7.1 LFs and Lexical Choices 92
7.2 LFs and Communicative Structure 94
7.3 LFs and Text Cohesion 95

Lexical Functions Across Languages **103**
Joseph E. Grimes
1 Introduction 103

2 Lexical Functions Are a Heuristic Tool 104
3 Why Do Lexical Functions Work? 107

Using Lexical Functions for the Extraction of Collocations from Dictionaries and Corpora **115**

Ulrich Heid

1 Introduction 115
2 Automatic Exploration of Language Resources 117
2.1 Analysis of Dictionary Articles 117
2.2 Analysis of Text Corpora 118
2.3 Collocation Discovery in Linguistic Resources: Dictionaries *vs.* Corpora 121
3 Exploiting LF Definitions for Discovery Procedures 122
3.1 Part of speech Combinations in Syntagmatic LFs . . . 123
3.2 Using the Definitions of **Oper**$_i$, **Func**$_i$, and **Labor**$_{ij}$ for Corpus Exploration 125
3.3 Exploiting Correlations between Semantic and Collocational Properties for Corpus Exploration Purposes 129
4 Extracting Collocations from Dictionaries 133
4.1 Problems of the Representation of Collocations in Dictionaries 134
4.2 An Analysis of a Few Collocationally Rich Dictionaries 135
4.3 Collocation Extraction from Definition Dictionaries . 137
4.4 Augmenting Lexical Descriptions with Information from Text Corpora 140
5 Summary 144

A Classification and Description of Lexical Functions for the Analysis of their Combinations **147**

Margarita Alonso Ramos & Agnès Tutin

1 Introduction 147
2 Classification of Lexical Functions 148
2.1 Brief Review of Previous Classifications of LFs 148
2.1.1 Zholkovsky and Mel'čuk's Classification . . 148
2.1.2 Steele and Meyer's Classification 149
2.1.3 Grimes' Classification 150
2.2 A New Classification of LFs 151

2.2.1 Paradigmatic LFs 152
2.2.2 Syntagmatic LFs 154
2.3 Justification of Our Classification 157
3 Descriptive Parameters 158
3.1 Semantic Parameters 158
3.2 Syntactic Categories 159
4 Combinations of Lexical Functions 160
4.1 Complex LFs 160
4.2 Compound LFs 161
4.3 LF Configurations 162
4.4 Internal Syntax of LF Combinations 162
4.4.1 Syntax of Complex LFs 163
4.4.2 Syntax of Compound LFs 163
4.4.3 Syntax of LF Configurations 164
5 Conclusion 164

A Case of Aspectual Polysemy, with Implications for Lexical Functions **169**

Alexander Nakhimovsky

1 Introduction 169
2 Telic and Atelic Readings 170
3 Atelic Perfectives in Russian 174
4 Aspectual Functions of $\mathbf{Oper_1}$ in English 177
5 Lexical Functions and Grammatical Meanings 178

On Dictionary Entries for Support Verbs: The Cases of Russian VESTI, PROVODIT′ and PROIZVODIT′ **181**

Tilmann Reuther

1 Focussing on the Problem 181
2 Examples from the TKS and the Deribas Glossary 183
2.1 TKS 183
2.2 Deribas Glossary 186
3 Semantic Links of $\mathbf{Oper_i}$-Verbs 187
3.1 VESTI 187
3.2 PROVODIT′/PROVESTI 189
3.3 PROIZVODIT′/PROIZVESTI 190

3.4 Support Verb Semantics: Step 1 190
4 Semantic Groups of Nouns in **Oper**-Collocations 192
4.1 VESTI 192
4.2 PROVODIT'/PROVESTI 193
4.3 PROIZVODIT'/PROIZVESTI 197
4.4 Support Verb Semantics: Step 2 198
5 Contextual Factors for the Choice of Support Verbs 199
5.1 Context Expressed by Adverbials of Time 199
5.1.1 Processual Meaning 199
5.1.2 Actual-Durative Meaning 200
5.2 Context Expressed by Adverbials of Manner and Instrument 200
5.2.1 Manner Adverbials 200
5.2.2 Instrumental Adverbials 201
6 Lexical Entries for the Support Verb VESTI in the ECD Format 201
7 Concluding Remarks 203

Lexical Functions and Lexical Inheritance for Emotion Lexemes in German **209**

Igor A. Mel'čuk & Leo Wanner

1 Introduction 209
1.1 The Statement of the Problem 209
1.2 The Data 211
1.3 The Methodology 212
1.4 The Structure of the Paper 213
2 Semantic and Syntactic Information in the ECD 214
2.1 Semantic Zone 214
2.2 Syntactic Zone 215
3 Emotion Lexemes in German 215
3.1 Semantics of Emotion Lexemes in German 216
3.1.1 The Notion of Semantic Dimension 216
3.1.2 List of Semantic Dimensions of Emotion Lexemes 219
3.2 Government Patterns of Emotion Lexemes in German 223
3.3 Restricted Lexical Co-occurrence of Emotion Lexemes in German 225

4 Towards a More Efficient Representation of Lexicographic Information 228
4.1 Discussion of Lexical Co-occurrence/Meaning Correlations 228
4.2 Implementing Syntactic Inheritance in an ECD 231
4.3 Implementing Lexical Inheritance Principle in an ECD 233
4.3.1 Co-occurrence Inheritance Technique in an ECD 233
4.4 Full *vs.* Compressed Lexical Entries: ANGST, HOFFNUNG, WUT 234
4.4.1 Three Full Lexical Entries of a German ECD 234
4.4.2 Three Compressed Lexical Entries of a German ECD 236
4.5 The Lexical Entry of GEFÜHL: the Generic Lexeme of the Semantic Field of Emotions 238
5 Conclusions 241

Some Procedural Problems in the Implementation of Lexical Functions for Text Generation 279

Lidija Iordanskaja, Myunghee Kim & Alain Polguère

1 Introduction 279
2 Use of Lexical Functions in the LFS/RTS Systems 281
3 General Principles behind the Implementation of Lexical Functions 282
4 Levels of Transition Where Lexical Functions Are Used . . . 284
4.1 SemR $\Longrightarrow$ DSyntR Transition 284
4.2 DSyntR $\Longrightarrow$ SSyntR Transition 286
5 Encoding the Values of Lexical Functions in the Lexicon . . 288
6 Implementation of Lexical Function Paraphrasing 289
7 LF Paraphrasing Rules Used in the LFS/RTS Systems 291
7.1 Six cases of RedSemR $\Longrightarrow$ DSyntR Transition Rules Using LFs 291
7.2 Criteria for Choosing Among Transition Rules 296
8 Conclusion 296

Generating Cohesive Text Using Lexical Functions **299**
Woongjae Lee & Martha Evens
1 Introduction 299
2 The Generation Process 300
3 The Choice of Referring Expressions 302
4 Generating Appropriate Collocations 304
5 The Lexical Database 306
6 Summary 306

RUSLO: **An Automatic System for Derivation in Russian** **307**
Natal'ia Percova
1 Introduction: Automatic System for Russian Derivation . . . 307
2 Linguistic Information in RUSLO 309
2.1 Formal Information 309
2.2 Semantic Information 311
3 Material for Further Research: Xlebnikov's Neologisms . . . 316

Bibliography **319**

Subject Index **336**

Name Index **353**

Introduction

Leo Wanner

University of Waterloo

The present volume contains a collection of articles that describe the recent work done on and with *Lexical Functions* (henceforth LFs) as put forward in the framework of the *Meaning-Text Theory*, MTT (Mel'čuk, 1974, 1988a). LFs are a means for a systematic description of what we would like to call "institutionalized" lexical relations. A lexical relation is institutionalized if it holds between two lexical units L_1 and L_2 and has the following characteristics: if L_1 is chosen to express a particular meaning M, its choice is predetermined by the relation of M to L_2 to such an extent that in case M and L_2 are given, the choice of L_1 is a language-specific automatism.[1] Often cited examples of institutionalized lexical relations are those between *aircraft* and *crew*, *sheep* and *flock*, *bachelor* and *confirmed*, *mountain* and *peak*, *influence* and *exert*, *attention* and *pay*, etc.

Contrastive studies of different instantiations of the same institutionalized relation are very illustrative. Thus, e.g., one *gives* somebody *credit*, but one *pays tribute to* somebody; one has a *good reputation*, but one has a *high opinion of* somebody. Further, the head of a *university* is called *president*, while the head of a *college* is the *principal*; a collection of *dogs* is called *pack*, while that of *cattle* is a *herd*; etc.

Institutionalized lexical relations are language-specific. In English you *take a walk*, and in French and German you 'make' it (*faire une promenade* and *einen Spaziergang machen*, respectively); in English you *deliver a speech*, in French you 'pronounce' or 'make' it (*prononcer* or *faire un discours*), in German and Italian you 'hold' it (*eine Rede halten* and *tenere un discorso*, respectively), and in Russian you 'pronounce' it (*proiznosit' reč'*). In German a collection of cattle is called 'herd' (*Herde*) as is a collection of sheep, while in English, Italian, and Russian different names are used: a

[1] In Coseriu's (1967) terminology, institutionalized lexical relations belong to the "norm of the language".