

CONTENTS

List of Figures	ix
List of Tables	xii
List of Contributors	xiii

PART I. INTRODUCTION AND OVERVIEW

1. Introduction	3
<i>A. Zampolli</i>	
2. Computational Approaches to the Lexicon: An Overview	17
<i>B. T. S. Atkins, Beth Levin, and A. Zampolli</i>	

PART II. THE COLLECTION AND PROCESSING OF TEXTUAL DATA

3. Not Chance but Choice: Criteria in Corpus Creation	49
<i>Gunnel Engwall</i>	
4. Encoding a Corpus in Machine-Readable Form: The Approach of the Text Encoding Initiative	83
<i>Stig Johansson</i>	
5. A Parser for Text Corpora	103
<i>Donald Hindle</i>	
6. Lexical Substitutability	153
<i>Kenneth Ward Church, William Gale, Patrick Hanks, Donald Hindle, and Rosamund Moon</i>	

PART III. THE THEORETICAL INFRASTRUCTURE OF LEXICAL ANALYSIS

7. Descriptive and Theoretical Syntax in the Lexicon	181
<i>Annie Zaenen and Elisabet Engdahl</i>	
8. Constructing Lexicon-Grammars	213
<i>Maurice Gross</i>	

9. Grammatical Data in the Lexicon <i>Eva Hajičová</i>	265
10. The Lexicon in a Computational Functional Grammar <i>Simon C. Dik</i>	279
11. A Richer Characterization of Dictionary Entries: The Role of Knowledge Representation <i>James Pustejovsky and Branimir Boguraev</i>	295
12. Lexicon Acquisition for NLP: A Consumer Report <i>Sergei Nirenburg</i>	313
13. Starting where the Dictionaries Stop: The Challenge of Corpus Lexicography <i>Charles J. Fillmore and B. T. S. Atkins</i>	349
PART IV. METHODOLOGY AND TOOLS	
14. A Methodology for the Construction of a Terminology Dictionary <i>Makoto Nagao</i>	397
15. The Lexicographical Workstation and the Scholarly Dictionary <i>Edmund Weiner</i>	413
16. A Lexical Workstation: From Textual Data to Structured Database <i>Nicoletta Calzolari and Eugenio Picchi</i>	439
Index	469

LIST OF FIGURES

3.1	The choice between depth and breadth	51
3.2	Selection procedures in corpus-building	52
3.3	Principal categories of 'parole'	56
4.1	Overall structure of the TEI-conformant text	87
4.2	File description	87
4.3	Encoding description	90
4.4	Profile description	90
5.1	A graphical representation of Fidditch output	109
5.2	The components of Fidditch	110
7.1	Levels of representation in GB	189
8.1	Distributional verbs versus support verbs	237
12.1	The architecture of the KBMT-89 system	319
12.2	A fragment of the concept network: text-group	320
12.3	Concept description: text-group	321
12.4	A fragment of the concept network: record-information	322
12.5	Concept description: record-information	323
12.6	An ONTOS screen	331
12.7	A prototypical comprehensive NLP system	335
12.8	Stand-alone knowledge-acquisition interfaces	339
12.9	Enhanced stand-alone knowledge-acquisition interfaces	341
12.10	An acquisition system coupled with an NLP system	343
13.1	RISK database: part of entry structure	365
13.2	The RISK scenario	366
13.3	RISK database: revised entry structure	368
14.1	KWIC index	402
14.2	A part of a term tree with relations	405
14.3	A model of term description and an example	408

15.1	The quotation slip: main elements	417
15.2	The quotation slip: file structure	419
15.3	The quotation slip: bibliographical and control information	420
16.1	The textual database (DBT) as a set of modules	442
16.2	The basic structure of the DBT	443
16.3	Text acquisition of the DBT	443
16.4	Main menu of the DBT system	444
16.5	Display of the subdivision of the text in its main logical parts	445
16.6	Display of the beginning of a logical section	446
16.7	Result of the query for a word with its frequency	446
16.8	Concordances of a word	447
16.9	User-defined context length	447
16.10	Concordance of a word with a user-defined length	448
16.11	Extension of the context to a whole logical part	448
16.12	Result of the query for words beginning with CONTR*	449
16.13	Result of the query for words ending with *(TURA/IONE)	449
16.14	Result of the query for patterns of letters within the words	450
16.15	User-defined family of words	451
16.16	Concordances of co-occurrences of two words	451
16.17	Selection of pivot and target languages in a multilingual corpus	451
16.18	Concordance of a word in the pivot language and its translations	452
16.19	Structure of the morphological processor	452
16.20	Structure of the generation program	453
16.21	Display of the generation of word forms	454
16.22	Template for the LDB lexical entry	455
16.23	Display of an entry with labelled fields	456
16.24	The word SCIENZA in the whole dictionary according to the different fields	456
16.25	Words beginning with SCIEN* in the whole dictionary	457
16.26	Contexts of a word belonging to the 'definition' field	458
16.27	Definition of a search family of words	458

16.28	Lemmas and contexts with co-occurrences of words	459
16.29	Search for hyponyms together with their definitions	459
16.30	Search for hyponyms	460
16.31	Search for lemmas whose definitions contain the relation AGENT	461
16.32	Search for lemmas whose definitions contain the relation SET_OF	461
16.33	Search for verbs in whose definition the word CIBO appears	462
16.34	Search for all the word forms of ANDARE in the corpus of texts	463
16.35	Search for all the hyponyms of ABITANTE in the corpus of texts	463
16.36	Display of left and right collocates of a key word	464

LIST OF TABLES

3.1	Principal features of the six categories	77
5.1	SVO triples extracted by Fidditch from an AP news story	118
5.2	What can you CURE?	124
5.3	Subjects and objects of JUROR	125
5.4	A sample of the verb-noun-preposition table	127
5.5	Verb-(NP)-preposition associations in Fidditch, AP sample, and Cobuild	129
6.1	Synonyms of ASK and ASK FOR	157
6.2	Mutual information: collocates of REQUESTED	160
6.3	A summary of words after REQUESTED	160
6.4	VO pairs: REQUEST	161
6.5	What do you typically REQUEST and what do you typically ASK FOR?	165
6.6	What do you typically DEMAND and what do you typically ASK FOR?	166
6.7	What do you typically DEMAND and what do you typically REQUEST?	167
6.8	Significant VO pairs observed and expected	170
6.9	Substitutability is not symmetric	171
6.10	Some asymmetric substitutions	172
8.1	Sample of a table of verbs with sentential subjects	221
8.2	A classification of frozen verbal expressions	231
8.3	A sample of a syntactic table of frozen verbal expressions whose structure is $N_0 V C_1$ (i.e. verbs with one direct frozen complement)	232
8.4	Classes of frozen adverbials	234
11.1	Selection by verbs: <i>play</i> versus <i>dictionary</i>	309
13.1	Treatment of verb RISK in three dictionaries	353
13.2	Treatment of verb RISK in ten dictionaries	354
13.3	Treatment of noun RISK in ten dictionaries	356
13.4	RISK verb sentence constituents as frame elements	369
13.5	Complements of verb RISK	373