Contents

The author and series editors xv
Foreword xvii
Introduction 1

1 Corpus creation

Introduction 13
Who should design a corpus? 13
A general corpus 13
Outline of corpus creation 14
- Electronic form 14
- Permissions 15

Design 15
- Spoken and written language 15
- Quasi-speech 16
- Formal and literary language 16
- Typicality 17
- Design criteria 17
- Period 18
- Overall size 18
- Sample size 19
- Whole documents 19
- Minimal criteria 20
- Provisional corpus 20
- Processing 20

Clean-text policy 21
Basic provision 22
Database 22
Maintenance 22
Different kinds of corpora 23
- Sample corpora 23
- Monitor corpora 24
- Features of a monitor corpus 25

2 Basic text processing

Introduction 27
Input 28
Words and word-forms 28
Text and vocabulary 29
Frequency list—first occurrence 30
Frequency list—alphabetical 31
Frequency list—frequency order 31
Word frequency profiles 32
Concordances 32
KWIC (Key Word in Context) 32
Longer environments 33
Ordering within concordances 33
Concordance processing 34
Text analysis statistics 34
Selective information 35
Intermediate categories 35
New approaches 36

3 The evidence of usage

Introduction 37
Existing descriptions 37
Native-speaker introspections 39
Language in use 39
Word-forms and lemmas 41
Concordances 42
Concordance evidence: an example 44
Sense 1: to refuse 47
Other senses 49

4 Sense and structure in lexis

Introduction 53
Evidence: main senses 53
Evidence: minor senses 55
Counter-examples: general 56
Counter-examples: first sense 57
Yield with an object 58
Yielding with an object 59
Yielded with an object 60
Descriptive and prescriptive study 60
Counter-examples: second and third senses 61
Yield as transitive verb 61
Doubtful cases 62
First minor sense 63
Conclusion 65

5 Words and phrases

Introduction 67
Phrasal verbs 67
Some numerical facts 68
Combinations of *set* + particle 69
The combination *set in* 70
Nouns 72
Verbs 72
Sense (i) 72
Sense (ii) 72
Minor senses 73
Sundry idioms 73
Set in as a phrasal verb 73
Word-forms 74
Subjects 74
A dictionary entry 75
Other phrasal verbs with *set* 75
Conclusion 78

6 The meeting of lexis and grammar

Introduction 81
- What is said about *of* 81
- A corpus view of *of* 82

Frequency 84

Of outside nominal groups 85

Of in nominal groups 85

Conventional measures 85

Less conventional measures 86

The status of headword 86

Focus nouns 87
- Focus on a part 87
- Focus on a specialized part 88
- Focus on a component, aspect, or attribute 88

Support 89
- Metaphor 90

Double-headed nominal groups 90
- Titles 90
- Nominalizations 91

Modification of first noun (N1) 93

Mopping up 94
- Superlative adjectives 94
- Fixed phrases 94
- Miscellaneous 94

Evaluation 95

Analysis of examples in Table 1 96
- Non-nominal instances of *of* 96
- Nominal group 96

Conclusion 98

7 Evaluating instances

Introduction 99
Throw away your evidence 99
Text and language 102
Meaning and structure 104
Procedure 105
Findings 107
Conclusion 108

8 Collocation

Introduction 109
Two models of interpretation 109
The open-choice principle 109
The idiom principle 110
Evidence from long texts 112
Collocation 115
Collocation of *back* 116
Analysis of the collocational pattern of *back* 117
Upward collocates: back 117
Downward collocates: back 118
Conclusion 121

9 Words about words

Introduction 123
Structure 124
Variation in co-text 126
About the word itself 126
What people mean 126
Structure: verb explanations 127
Animate subjects 127
Inanimate subjects 129
Mixed subjects 130
Operators 130
Summary 132

Analysis of the second part 132
First chunk 132
Second chunk 134
Discussion 134
Applications 135
Inferences and implications 135
Conclusion 136
Summing up 137

Appendix I 139
Appendix II 150
Appendix III 154
Bibliography 165
Glossary 169
Index 177

Index

Entries relate to the Introduction, Chapters 1 to 9, and the glossary. References to the glossary are indicated by 'g' after the page number.

adjectives, superlative, with *of* 94
alphabetical order,
  concordances 33–4, 147
  frequency lists 31, 141, 142
ambiguity 104–5
animate subjects, explanation of verbs 127–9
automatic parsing 36

back,
  collocation 116–21
  meanings 112
'bank of English' 26
Biber, D. 18
Birmingham Collection of English Texts *now* The Bank of English 24, 40, 53–4
Brown University corpus 19, 23–4, 40, 99–100

CALL (Computer Assisted Language Learning) 36
categories, categorization 35–6, 49
character(s) 28, 169g
Chomsky, N 103
chunking 124, 132–4
citation(s) 40, 169g
  in concordances 43–4
  forms 105, 108
clean-text policy, 21–2
Cobuild 1–3, 169g
  dictionary 2, 75, 123–4, 127
  English grammar 89
  frequency count 143
coding standards 21
coherence 102
Collins English Dictionary (*CED*),
  back 112
  decline 44–50
  pursue 113
collocate(s), collocation 3, 71, 105–7, 109–21, 170g
comment, part of sentence 124, 125
competence vs. performance 103
Comprehensive Grammar of the English Language, A (CGEL) 100
Computer Assisted Language Learning (CALL) 36
computer-held corpora 14–15
computer processing 27–36
concordance(s) 40–41, 42–51, 105–7, 170–1g
context 171g
copyright permission 15
'core' meaning 113
corpus/corpora 171–2g
  creation 13–26
  kinds 23–6
  size 18–19, 100
corrections to corpora 22–3
co-text 124, 125, 126–7, 172g
counter-examples 99

Daley, R. 106
databases from corpora 22
data processing 27–36
decline 44–51, 150–3
delexicalization 113
descriptive vs. prescriptive study 60–61
design of corpora 15–21
deviance, text and language 102
dictionaries 7, 38–9, 75, 134–5
discourse 172g
documents in corpora 19
double-headed nominal groups 90–93
downward collocation 115–16, 118

electronic form of corpus 14–15
encoding standards 21
evaluation of instances 99–108
evidence of usage 37–51
examples, citing and discarding 99, 100
existing descriptions 37–9
extracts in corpora 23

first order word list 140
fixed phrases with *of* 94
focus nouns with *of* 87–9

formal language 16
framework of explanation 132–3
Francis, W. 23, 99
frequency 18–19, 30–32, 142, 144

genres, classification in corpora 19, 23
gloss 132–3
grammar and lexis 81–88

Hanks, P.W. 123
headword,
  similarity to lemma 42
  status 86–7
Hofland, K. 19, 40, 99
homographs 42

idiom 8, 172–3g
idiom principle 110–15, 173g
implications from explanations 135–6
inanimate subjects, explanations of verbs 129–30
inferences from explanations 135–6
informal language 16
informant testing 39
input of text 14–15, 28
instances,
  citation 40
  evaluation 99–108
interpretation 109–15
introspections, native-speaker 39
is, concordance 146–9

Johansson, S. 19, 23, 40, 99
Johnson, Dr 40
Jones, S. 24, 27, 106, 117

Kučera, H. 23, 99
KWIC (Key Word in Context) 32–3, 43, 146–8, 173g

Lancaster-Oslo-Bergen (LOB) corpus 19, 23, 40, 99–100
language vs. text 102–4
langue vs. *parole* 103
legal aspects, copyright 15
lemma(s), lemmatization 8, 41–2, 173–4g
lexicographical constructs 38
lexicography 40, 123
lexis 174g
  and grammar 81–98
  sense and structure 53–65
  and syntax 103–4
literary language 16–17
LOB (Lancaster-Oslo-Bergen) corpus 19, 23, 40, 99–100
London-Lund corpus 24

maintenance of corpora 22–3
Martinet, A.V. 82
meaning,
  'core' 113
  explanation of 123–37
  and structure 104–6
  units 6–9
metaphor, with *of* 90
Mindt, D. 3
mixed subjects, explanation of verbs 130
Modern English Grammar, A 100
modification of nouns 93–4
monitor corpora 9, 24–6

native-speaker introspections 39
natural language 5–6
naturalness 5–6, 174–5g
neutral collocates 116
node(s) 105–6, 115, 175g
nominal groups 85–98
nominalizations 91–3

occurrences, frequency 30–31, 142
OED (*Oxford English Dictionary*) 40
of 81–98
of course, idiom principle 110–11
open-choice principle 109–10, 175g
operators 124, 125, 130, 131, 132, 133
optical scanning for input 14–15
ordering, in concordances 33–4, 43, 140–2, 146, 147, 148
Oxford English Dictionary (*OED*) 40

paraphrase 136–7
parole vs. *langue* 103
performance vs. competence 103
performative use 47
period of time, coverage of corpora 18
permissions, copyright 15
Phillips, M. 3, 31, 45
phrasal verbs 67–79
phrase-meaning vs. word-meaning 71
phrases, with *of* 94
pragmatics 3
prescriptive vs. descriptive study 60–61
processing of corpora and texts 20–21, 27–36
Productive Features 61
provisional corpora 20

quasi-speech 16
Quirk, R. 24, 82

Reed, A. 27
register 110
Renouf, A. 3, 13, 17, 40, 54
reverse alphabetization, in
concordances 34, 148
running words 30, 175g

sample corpora 19, 23–4
Saussure, F. de 103
scanning for input 14–15
second, meanings 106–7
selection,
for concordances 35, 43
for corpora 20, 40–41
semantics 3
sense(s),
and frequency 101–2
and structure 53–65
and syntax, co-ordination 154–6
sentences, structure 124–5
set, phrasal verbs 67–79
Sinclair, J.M. 2, 3, 5, 20, 24, 27, 38, 61, 68, 89, 106, 108, 109, 117, 123
slot-and-filler model 109–10
span 175g
spoken language 15–16
Stock, P.F. 49
structure,
and meaning 104–6
and sense 53–65
statements 124–5
verb explanations 127–32
subjects, in explanations of verbs 127–30
substitutability 82
superlative adjectives, with *of* 94
supporting nouns, with *of* 89–94
Survey of English Usage 100
Svartvik, J. 24
syntax,
and lexis 103–4
and sense, co-ordination 154–6

text analysis statistics 34–5, 149
Text Encoding Initiative 21
text(s) 175–6g
input 14–15
processing 20–21, 27–36
representation, standards 21
selection 15–17
vs. language 102–4
'text-transitive' 49
The Bank of English *formerly* Birmingham Collection of English Texts 24, 40, 53–4
the, frequency 18, 143
thesauri 7
Thomson, A.J. 82
time indicators, coverage of corpora 18
titles, double-headed nominal groups 90–91
topic, part of sentence 124, 125
transitivity 49
typicality 17

units of meaning 6–9
upward collocation 116, 117
usage, evidence of 37–51

verb explanations, structure 127–32
verbs, phrasal 67–79
vocabulary 29–30, 176g
clustering 31

Willis, J.D. 3
word-form(s) 28–9, 176g
frequency *see* word(s), frequency
and lemmas 41–2
see also concordances; vocabulary
word-meaning vs. phrase-meaning 71
word(s) 28–9, 176g
explanation 123–37
frequency 18–19, 30–32, 142, 144
and phrases 67–79
written language 15–16

yield 53–65, 157–63

Zipf, G.K. 18