CORPUS LINGUISTICS
Critical Concepts in Linguistics

Edited by
Wolfgang Teubert
&
Ramesh Krishnamurthy

Volume I
# CONTENTS

*Acknowledgements*  
xix

*Chronological table of reprinted articles and chapters*  
xxiii

*Foreword*  
xxxvii

General introduction  
1

## VOLUME I

### PART 1

**Theoretical aspects of corpus linguistics**  
39

1. *The importance of corpus linguistics to understanding the nature of language*  
Wallace Chafe  
41

2. *Does corpus linguistics exist? Some old and new issues*  
Jan Aarts  
58

3. *The corpus-driven approach*  
Elena Tognini-Bonelli  
74

4. *“Corporate bridges” twixt text and language*  
Robert de Beaugrande  
93

5. *Corpora, databases and the organisation of linguistic data*  
Gerry Knowles  
119

6. *Writing, hermeneutics, and corpus linguistics*  
Wolfgang Teubert  
134

7. *Apples and oranges: on comparing data from different corpora*  
Hans Lindquist and Magnus Levin  
160
## CONTENTS

### PART 1

8 Corpus linguistics, Chomsky and fuzzy tree fragments  
  BAS AARTS  

9 Meaning in the framework of corpus linguistics  
  JOHN SINCLAIR  

10 “Corpus linguistics” or “computer-aided armchair linguistics”  
  CHARLES FILLMORE  

### PART 2

#### History of corpus linguistics

11 English lexical collocations: a study in computational linguistics  
  JOHN SINCLAIR AND SUSAN JONES  

12 Interview with John Sinclair conducted by Wolfgang Teubert  
  JOHN SINCLAIR AND WOLFGANG TEUBERT  

13 Problems of assembling and computerizing large corpora  
  W. NELSON FRANCIS  

14 Towards a description of English usage  
  RANDOLPH QUIRK  

15 The value of a corpus in English language research: a reappraisal  
  GEOFFREY LEECH  

16 Claimed and unclaimed sources of corpus linguistics  
  JACQUELINE LÉON  

### VOLUME II

#### Acknowledgements

ix  

#### PART 3

#### Corpus composition and compilation

17 Corpora  
  GEOFFREY LEECH
CONTENTS

18 Compiling and using the IJS-ELAN parallel corpus
TOMAZ ERJAVEC

19 Where did we go wrong? A retrospective look at the
British National Corpus
LOU BURNARD

20 Towards an English–Norwegian parallel corpus
STIG JOHANSSON AND KNUT HOFLAND

21 The American National Corpus: overall goals and the
first release
RANDI REPPEN AND NANCY IDE

22 Parallel corpora: a real-time approach to the study of
language change in progress
CHRISTIAN MAIR

23 The development of the International Corpus of English
SIDNEY GREENBAUM

24 Corpus design criteria
SUE ATKINS, JEREMY CLEAR AND NICHOLAS OSTLER

25 Representativeness in corpus design
DOUGLAS BIBER

26 The computer learner corpus: a versatile new source of data
for SLA research
SYLVIANE GRANGER

27 Evolution and present situation of corpus research in China
ZHIWEI FENG

PART 4
Standardisation, alignment, tagging and corpus related
software

28 The detection of inconsistency in manually tagged text
HANS VAN HALTEREN

29 Comparing corpora
ADAM KILGARRIFF

30 From the Rosetta stone to the information society: a survey
of parallel text processing
JEAN VÉRONIS
<table>
<thead>
<tr>
<th>Article</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Char_align: a program for aligning parallel texts at the character level</td>
<td>290</td>
</tr>
<tr>
<td>KEN CHURCH</td>
<td></td>
</tr>
<tr>
<td>PC analysis of key words – and key key words</td>
<td>303</td>
</tr>
<tr>
<td>MIKE SCOTT</td>
<td></td>
</tr>
<tr>
<td>AUTASYS: grammatical tagging and cross-tagset mapping</td>
<td>318</td>
</tr>
<tr>
<td>ALEX C. FANG</td>
<td></td>
</tr>
<tr>
<td>Meaning and interpretation of markup</td>
<td>334</td>
</tr>
<tr>
<td>MICHAEL SPERBERG-McQUEEN, CLAUS HUITFELDT AND ALLEN RENEAR</td>
<td></td>
</tr>
<tr>
<td>A formal framework for linguistic annotation</td>
<td>355</td>
</tr>
<tr>
<td>STEVEN BIRD AND MARK LIBERMAN</td>
<td></td>
</tr>
<tr>
<td>A framework of a mechanical translation between Japanese and English by analogy principle</td>
<td>407</td>
</tr>
<tr>
<td>MAKOTO NAGAO</td>
<td></td>
</tr>
<tr>
<td>Intuition and annotation: the discussion continues</td>
<td>415</td>
</tr>
<tr>
<td>JOHN SINCLAIR</td>
<td></td>
</tr>
<tr>
<td>From Firth principles: computational tools for the study of collocation</td>
<td>436</td>
</tr>
<tr>
<td>JEREMY CLEAR</td>
<td></td>
</tr>
<tr>
<td>Bi- and trilingual alignment and concordancing as machine aids to human translation</td>
<td>457</td>
</tr>
<tr>
<td>MICHAEL P. OAKES AND BARBARA LEWANDOWSKA-TOMASZCZYK</td>
<td></td>
</tr>
<tr>
<td>One sense per discourse</td>
<td>467</td>
</tr>
<tr>
<td>WILLIAM GALE, KENNETH CHURCH AND DAVID YAROWSKY</td>
<td></td>
</tr>
</tbody>
</table>
CONTENTS

VOLUME III

Acknowledgements vii

PART 5
Lexicography, collocation, idioms and phraseology 1

41 The search for units of meaning
JOHN SINCLAIR 3

42 Irony in the text or insincerity in the writer?
The diagnostic potential of semantic prosodies 30
BILL LOUW

43 Semantic prosodies in English and Portuguese:
A contrastive study 51
TONY BERBER SARDINHA

44 Metaphorical polysemy and paradigmatic relations:
a corpus study 67
ALICE DEIGNAN

45 Patterns of word usage in corpus linguistics 85
RŪTA MARCINKEVIČIENĖ

46 The weight of words: an investigation of lexical gravity 97
OLIVER MASON

47 Corpus-based lexicon building: an overview across projects,
problems, approaches 112
NICOLETTA CALZOLARI

48 Enthusiasm and condescension: changing norms of
meaning and use 140
PATRICK HANKS

49 The distribution of idioms in English 154
ROSAMUND MOON

50 Collocations and semantic profiles: on the cause of
the trouble with quantitative studies 166
MICHAEL STUBBS

51 Asian or Western realities? Collocations in
Singaporean-Malaysian English 194
VINCENT OOI
CONTENTS

52 Lexis and lexicographers: the vocabulary of dictionary definitions
GEOFF BARNBROOK 213

53 On the phraseology of spoken English: the evidence of recurrent word-combinations
BENGT ALTENBERG 224

54 Textual colligation: a special kind of lexical priming
MICHAEL HOEY 245

55 Lexicographic relevance: selecting information from corpus evidence
SUE ATKINS, CHARLES FILLMORE AND CHRISTOPHER JOHNSON 269

PART 6
Terminology 301

56 Towards a corpus-based approach to terminography
LYNNE BOWKER 303

57 In search of representativity in specialised corpora: categorisation through collocation
GEOFFREY WILLIAMS 325

58 Making a workable glossary out of a specialised corpus: term extraction and expert knowledge
CHRISTINE CHODKIEWICZ, DIDIER BOURI GAULT AND JOHN HUMBLEY 346

59 Comment accéder aux éléments définitoires dans les textes spécialisés?
JENNIFER PEARSON 365
CONTENTS

VOLUME IV

Acknowledgements ix

PART 7

Grammar 1

60 Looking at looking: functions and contexts of progressives in spoken English and 'school' English 3
UTE RÖMER

61 The textlinguistic dimension of corpus linguistics: the support function of English general nouns and its theoretical implications 15
MICHAELA MAHLBERG

62 Between and through: the company they keep and the functions they serve 26
GRAEME KENNEDY

63 What is a grammatical rule? 41
DIETER MINDT

64 Will corpus linguistics revolutionize grammar teaching in the 21st century? 55
SUSAN CONRAD

65 Why grammar is beyond belief 67
MICHAEL HOEY

66 A corpus-driven approach to grammar: principles, methods and examples 84
GILL FRANCIS

67 A quantitative study of polarity and primary tense in the English finite clause 104
MICHAEL A. K. HALLIDAY AND ZOE JAMES

68 Subordinate clauses in English 138
JOHN KIRK

69 Lexical gaps 154
GÖRAN KJELLMER
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>A preliminary analysis of collocational differences in monolingual comparable corpora</td>
<td>366</td>
</tr>
<tr>
<td></td>
<td>MARCO BARONI AND SILVIA BERNARDINI</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>A corpus-based view of similarity and difference in translation</td>
<td>384</td>
</tr>
<tr>
<td></td>
<td>MONA BAKER</td>
<td></td>
</tr>
</tbody>
</table>

### VOLUME V

#### Acknowledgements

vii

#### PART 9

Critical discourse analysis / evaluation / stylistics / rhetoric

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>84</td>
<td>Speaking and writing in the university: a multidimensional comparison</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>DOUGLAS BIBER, SUSAN CONRAD, RANDI REPPEN, PAT BYRD AND MARIE HELT</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>Description and interpretation in critical discourse analysis</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>JOHN FLOWERDEW</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>Corpus evidence on Australian style and usage</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>PAM PETERS</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>Corpus, comparison, culture: doing the same things differently in different cultures</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>GEOFF THOMPSON</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>Footing shift for attribution: ‘According to the New York Times this morning...’</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>ALAN PARTINGTON</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>Modal verbs in academic writing</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>PAUL THOMPSON</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>Compressed noun-phrase structures in newspaper discourse: the competing demands of popularization vs. economy</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>DOUGLAS BIBER</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>A province of a federal superstate ruled by an unelected bureaucracy: keywords of the Euro-sceptic discourse in Britain</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>WOLFGANG TEUBERT</td>
<td></td>
</tr>
</tbody>
</table>
92 Ethnic, racial and tribal: the language of racism?  RAMESH KRISHNAMURTHY  179

93 On the use of corpora in the analysis of forensic texts  MALCOLM COULTHARD  201

94 Domestic discord, rocky relationships: semantic prosodies in representations of marital violence in the O.J. Simpson trial  JANET COTTERILL  219

95 Corpus-based analysis of evaluative lexis  JOANNA CHANNELL  244

PART 10
Language history / historical linguistics  265

96 Gender differences in the evolution of standard English: evidence from the Corpus of Early English Correspondence  TERTTU NEVALAINEN  267

97 Politeness and modal meaning in the construction of humiliative discourse in an early eighteenth-century network of patron-client relationships  SUSAN FITZMAURICE  288

98 Grammaticalisation from side to side: on the development of beside(s)  MATTI RISSANEN  318

99 Advanced research on syntactic and semantic change with the Corpus del Español  MARK DAVIES  337

100 Collocational and idiomatic aspects of verbs in Early Modern English: a corpus-based study of MAKE, HAVE, GIVE, TAKE, and DO  MERJA KYTÖ  349

101 Three changing patterns of verb complementation in Late Modern English: a real-time study based on matching text corpora  CHRISTIAN MAIR  386
CONTENTS

VOLUME VI

Acknowledgements vii

PART 11

Language teaching 1

102 Should you be persuaded: two samples of data-driven learning materials 3
TIM JOHNS

103 The pedagogical value of native and learner corpora in EFL grammar teaching 23
FANNY MEUNIER

104 A bird’s-eye view of learner corpus research 44
SYLVIANE GRANGER

105 Cross-disciplinary comparisons of hedging: some findings from the Michigan Corpus of Academic Spoken English 73
DEANNA POOS AND RITA SIMPSON

106 English writing by Dutch-speaking students 92
PIETER DE HAAN

107 Corpora in the classroom: an overview and some reflections on future developments 102
SILVIA BERNARDINI

108 A new academic word list 123
AVERIL COXHEAD

109 A corpus-based approach to Modern Greek language research and teaching 150
DIONYSIS GOUTSOS, OURANIA HATZIDAKI AND PHILIP KING

110 Norms for the Indian English classroom: a corpus-linguistic perspective 157
JOYBRATO MUKHERJEE
PART 12
Spoken language / discourse studies 175

111 Integrated and fragmented worlds: EAP materials and corpus linguistics 177
JOHN SWALES

112 Engaged listenership in spoken academic discourse: the case of student–tutor meetings 196
FIONA FARR

113 The corpus of English as lingua franca in academic settings 217
ANNA MAURANEN

114 The performance of speech acts in workplace conversations and the teaching of communicative functions 232
ALMUT KOESTER

115 Common language: corpus, creativity and cognition 251
RONALD CARTER

116 Functional grammar and discourse studies 276
CHARLES MEYER

117 Talking back: “Small” interactional response tokens in everyday conversation 292
MICHAEL McCARTHY

118 Spoken corpus design 320
STEVE CROWDY

119 The spoken language corpus: a foundation for grammatical theory 334
MICHAEL A. K. HALLIDAY

Index 363

xviii