

Quantitative Linguistics 64

Editors

Reinhard Köhler
Gabriel Altmann
Peter Grzybek

Mouton de Gruyter
Berlin · New York

Word Frequency Studies

by

Ioan-Iovitz Popescu

in co-operation with

Gabriel Altmann

Peter Grzybek

Bijapur D. Jayaram

Reinhard Köhler

Viktor Krupa

Ján Mačutek

Regina Pustet

Ludmilla Uhlířová

Matummal N. Vidya

Mouton de Gruyter

Berlin · New York

Contents

Preface	v
1 Introduction	1
2 Problems and presentations	5
2.1 Problems	5
2.2 Presentations	9
3 The h- and related points	17
3.1 The h -point	17
3.1.1 A first look at vocabulary richness	29
3.2 The k -point	35
3.2.1 Definition	35
3.2.2 A second look at vocabulary richness	36
3.2.3 Indicator b	44
3.3 The m -point	48
3.3.1 The m -coverage of text and vocabulary richness	52
3.4 Gini's coefficient and the n -point	54
3.4.1 The n -point	63
3.5 The role of N and V	70
4 The geometry of word frequencies	73
4.1 Introduction	73
4.2 The rank frequency distribution	75
4.3 The spectrum	81
5 The dynamics of word classes	87
6 Thematic concentration of the text	95
7 Crowding, pace filling and compactness	101
7.1 Crowding	101
7.2 Pace filling	103
7.3 Compactness	107

8	Autosemantic text structure	111
8.1	Introduction	111
8.2	The probability of co-occurrence	113
8.3	The construction of a graph	121
8.4	Degrees	124
9	Distribution models	127
9.1	General theory	127
9.2	Special cases	130
9.3	Applications	133
9.4	The spectrum	143
9.5	Evaluations	152
9.6	Ord's criterion	154
9.7	Repeat rate and entropy	165
9.8	Word classes	185
10	The relation of frequency to other word properties	195
11	Word frequency and position in sentence	203
11.1	Introduction	203
11.2	Runs of binary data	206
11.3	Runs of multiple data	209
11.4	Absolute positions	210
11.4.1	Word frequency in the final position	211
11.4.2	Word frequency in the initial position	213
11.5	Relative position	214
11.6	Frequency motifs	218
11.7	Distances between hapax legomena	227
12	The type-token relation	231
12.1	Introduction	231
12.2	Standard measurement	234
12.3	Köhler-Galle method	239
12.4	The ratio method	240
12.5	Stratified measurement (the window method)	241
12.6	The TTR of F -motifs	244
13	Conclusions	249

14 Appendix: Texts	253
References	265
Subject index	271
Author index	275