Studies
in
Quantitative Linguistics
12

Radek Čech
Gabriel Altmann

Problems in Quantitative Linguistics 3

RAM - Verlag

Problems in Quantitative Linguistics 3

by

Radek Čech
Gabriel Altmann

Dedicated to Reinhard Köhler on the occasion of his 60th birthday

2011
RAM-Verlag

Studies in quantitative linguistics

Editors
Fengxiang Fan (fanfengxiang@yahoo.com)
Emmerich Kelih (emmerich.kelih@uni-graz.at)
Reinhard Köhler (koehler@uni-trier.de)
Ján Mačutek (jmacutek@yahoo.com)
Eric S. Wheeler (wheeler@ericwheeler.ca)

1. U. Strauss, F. Fan, G. Altmann, *Problems in quantitative linguistics 1*. 2008, VIII + 134 pp.
2. V. Altmann, G. Altmann, *Anleitung zu quantitativen Textanalysen. Methoden und Anwendungen.* 2008, IV+193 pp.
3. I.-I. Popescu, J. Mačutek, G. Altmann, *Aspects of word frequencies*. 2009, IV +198 pp.
4. R. Köhler, G. Altmann, *Problems in quantitative linguistics 2.* 2009, VII + 142 pp.
5. R. Köhler (ed.), *Issues in Quantitative Linguistics.* 2009, VI + 205 pp.
6. A. Tuzzi, I.-I. Popescu, G.Altmann, *Quantitative aspects of Italian texts*. 2010, IV+161 pp.
7. F. Fan, Y. Deng, *Quantitative linguistic computing with Perl.* 2010, VIII + 205 pp.
8. I.-I. Popescu et al., *Vectors and codes of text*. 2010, III + 162 pp.
9. F. Fan, *Data processing and management for quantitative linguistics with Foxpro.* 2010, V + 233 pp.
10. I.-I. Popescu, R. Čech, G. Altmann, *The lambda-structure of texts*. 2011, II + 181 pp
11. E. Kelih et al. (eds.), *Issues in Quantitative Linguistics Vol. 2*. 2011, IV + 188 pp.
12. R. Čech, G. Altmann, *Problems in quantitative linguistics 3*. 2011, VI + 168 pp.

ISBN: 978-3-842303-08-8

© Copyright 2011 by RAM-Verlag, D-58515 Lüdenscheid

RAM-Verlag
Stüttinghauser Ringstr. 44
D-58515 Lüdenscheid
RAM-Verlag@t-online.de
http://ram-verlag.de

Preface

The third volume of *Problems* has the same structure and the same aim as the first two volumes. The problems can be used for writing essays for final examinations, dissertations or, in European context, "habilitations". The authors would be very pleased if solutions to some problems would be published in some appropriate journals. They also hope that each problem presented here will be developed further both in empirical and theoretical directions. In many cases only suggestions and hints are given but the authors hope that the reader will go his own way and systematize the given problem, i.e. subsume it under a more general problem, link it with other phenomena and, at last, propose a theoretical derivation with linguistic substantiation. However, even empirical testing using other texts/languages and empirical generalizations would be of great value.

Each problem automatically evokes other problems which are either collateral or hierarchic, i.e. concern problems at the same linguistic level or at a different one. As a matter of fact, local problems can be solved in isolation adequately or ad hoc, but in theoretical research the only criterion of a useful solution is systematization. Hence we recommend the reader not to stop at the solution of a problem for a single text or text sort or author or language but to extend the analysis at least by means of empirical generalization to different objects. Since the readers are linguists, the extension of the scope of study to other languages may not be a serious problem.

In the present volume the problems are classified into six groups: phonology and script, grammar, semantics, synergetics, text analysis and a group of mixed problems. If there is a hint to a similar problem in the previous volumes, it is recommended to solve first the simpler task and generalize the problem step by step.

The mathematics necessary for solutions is simple and can either be found in text-books of statistics, or, if not, then the reader finds instructions directly in the "Procedure" accompanying each problem. It is recommended to read at least some of the works given in the References. Many times we quote old works in order to show that the problem itself is nothing new but obtains a different look in its new quantitative costume.

Acknowledgments
We are obliged to Eric Wheeler who read the book and helped us to improve our style. Radek Čech was supported by the Czech Science Foundation, grant no. 405/08/P157.

Radek Čech
Gabriel Altmann

Contents

Preface I

1. Phonology and script 1

1.1. Phoneme distribution: structural gaps 1
1.2. Stress placement in isosyllabic poems 3
1.3. Vowel duration 5
1.4. Homophones 6
1.5. Phonetic difference and frequency 8
1.6. Script motifs 12
1.7. Formal problems of rongorongo 14
1.8. Symbol frequencies 15

2. Grammar 25

2.1. Word form diversification
2.2. Diversification of case in Finnish 26
2.3. Inflection and derivation 28
2.4. Transitivity 29
2.5. Morph length 31
2.6. Distribution of morphological features 33
2.7. Paradigmatic expansion of words 34
2.8. Morpheme dynamics 35
2.9. Cohesion of compounds 37
2.10. Symmetry of sentence structure 38
2.11. Transitivity as a text parameter 40
2.12. Transitivity and text deployment 44
2.13. Transitivity and the verb 45
2.14. Transitivity and child development 46
2.15. Transitivity and aspect 47
2.16. The distribution of full valency frames 49
2.17. Syntactic network analysis: hub/authority weight versus degrees 52
2.18. Syntactic network analysis: hub/authority weight as a text parameter 53
2.19. Proper names as a high Transitivity feature 54
2.20. Proper names and aspect of verbs 58

3. Semantics 61

3.1. Yesypenko's linguistic world view 61
3.2. Polysemy and synonymy 62
3.3. The role of synonymy in self-regulation 64
3.4. Distribution of semantic roles and frames (corpus) 66

3.5. Verb classes 67
3.6. Verb classes and valency 70
3.7. Meaning specificity and compounding 70

4. Synergetics 72

4.1. Arens'law 72
4.2. Frequency and polytexty 73
4.3. Frequency and noun generality 75
4.4. Increase of word length 76
4.5. Text length vs. word length 76
4.6. Word length and phoneme inventory 77
4.7. Word length distributions 79
4.8. Full valency and the frequency of verbs 84
4.9. Full valency and the length of verbs 85
4.10. Full valency and polysemy 86
4.11. Full valency and synonymy 87
4.12. Full valency and compounding 88
4.13. Full valency and derivation 89
4.14. Full valency and Menzerath's law 89
4.15. Syntactic network analysis 90
4.16. Typology 93

5. Text analysis 96

5.1. Text development 96
5.2. Cumulative text development 97
5.3. Experiments with arc length 98
5.4. The binary syntactic code of sentence 100
5.5. The binary code of text 104
5.6. The development of writer's view 105
5.7. Hřebíček's hypothesis 106
5.8. Chaining and distances 107
5.9. Sonnet 1 109
5.10. Sonnet 2 110
5.11. Text activity 1 113
5.12. Text activity 2 115
5.13. Assonance 116
5.14. Frequency and position of words in a sentence 117
5.15. Word length and position in a sentence 118
5.16. Development of rhyme 120
5.17. Vowel auto-affinity in poems 121
5.18. The lambda indicator and Ord's criterion 122
5.19. The lambda-indicator and Busemann's adjective-verb ratio 124
5.20. Sentence length in discourse 125
5.21. Sentence length sequences in discourse 126

5.22. Properties of illocutive graphs 127
5.23. The frequency sequence of words in text 129
5.24. Frequency motifs in text 130
5.25. Thematic concentration 131
5.26. Dialogue: thematic concentration of participants 133
5.27. Text compactness 134
5.28. Frequency indicator A 135
5.29. A vocabulary richness indicator 136
5.30. Vocabulary richness and lambda-structure 137
5.31. Gini's coefficient and vocabulary richness 138
5.32. Vocabulary richness project 139
5.33. A textological project 141
5.34. Text diffusity 142
5.35. Aggregation in poetry 143

6. Different issues 146

6.1. Rank-frequency distribution of classifications 146
6.2. Ranking and classification of verbs 147
6.3. Sentence length development in German 149
6.4. Verb development 151
6.5. History 151
6.6. Psychoanalytic word categories 153
6.7. Language of children 155
6.8. Vocalic language 156

Author index 158

Subject index 165