

**Problems  
in  
Quantitative Linguistics  
4**

**by**

**Reinhard Köhler  
Gabriel Altmann**

**2014  
RAM-Verlag**

## Studies in quantitative linguistics

### Editors

Fengxiang Fan ([fanfengxiang@yahoo.com](mailto:fanfengxiang@yahoo.com))  
Emmerich Kelih ([emmerich.kelih@uni-graz.at](mailto:emmerich.kelih@uni-graz.at))  
Reinhard Köhler ([koehler@uni-trier.de](mailto:koehler@uni-trier.de))  
Ján Mačutek ([jmacutek@yahoo.com](mailto:jmacutek@yahoo.com))  
Eric S. Wheeler ([wheeler@ericwheeler.ca](mailto: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.
13. R. Köhler, G. Altmann (eds.), *Issues in Quantitative Linguistics Vol 3*. 2013, IV + 403 pp.

ISBN: 978-3-942303-22-4

© Copyright 2014 by RAM-Verlag, D-58515 Lüdenscheid

RAM-Verlag  
Stüttinghauser Ringstr. 44  
D-58515 Lüdenscheid  
[RAM-Verlag@t-online.de](mailto:RAM-Verlag@t-online.de)  
<http://ram-verlag.de>

## Preface

The fourth volume of *Problems in Quantitative Linguistics* provides once more new evidence for the truth of the statement that research will never run out of challenges. Any aspect of a unit, a property, a level of linguistic analysis, a language, an individual text, a text sort, etc. can be specified and deepened or generalized. New ways of looking at language are found, new methods are developed and new questions are asked. Established units and well-known properties are connected in new ways and incorporated in control cycles which furnish us new hypotheses and even theories. Besides, new vistas taken from other sciences can be introduced and the linguistic reality can be seen as something that has analogies with the “rest” of the world. Today, we use a great part of “quantitative” mathematics, see the language from psychological, communicative, pragmatic, social, grammatical, textological, evolutionary, diversificational, stochastic points of view and introduce ever further views like that of systems theory, graph theory, fractals, time dependent processes, etc.

Testing the presented hypotheses does not only concern their corroboration but rather the search for more general hypotheses, or, more specific ones with inclusion of some boundary conditions.

The present volume contains again diversified problems which can be used for writing contributions to journals, dissertations or for organizing projects in quantitative linguistics. There are no exercises in the book, but problems whose solution would contribute to the development of this science. The readers are invited to write articles and send them to the journals *Glottometrics*, *Glottology* or *Journal of Quantitative Linguistics*.

We are aware of the fact that some problems represent complex projects. Do not try, in these cases, to solve all details at once and set up a complex theory at the first attempt and in a single step. Solve only a first, partial problem, collect data from many texts or languages; then solve the second partial aspect of the problem and generalize stepwise. The linguistic aspects and data collection must be made by a linguist (not by a mathematician), the mathematician should help solving the mathematical problems. A programmer can be consulted only if the linguist is able to present all definitions in a formal way.

In the present volume there is more syntax than in the previous ones and many problems are more complex. The authors are ready to help researchers who are interested in this kind of investigations.

Reinhard Köhler  
Gabriel Altmann

# Contents

## Preface

<b>1. Syntax</b>	<b>1</b>
1.1. Event integration	1
1.2. Cline of grammaticality	2
1.3. Complementation scale and co-lexicalization	3
1.4. Complementation scale and subordinating morphemes	4
1.5. Voice diversification	5
1.6. Remote referent	7
1.7. Cohesion, coherence and continuity	8
1.8. Anaphoric distance	9
1.9. Cataphoric persistence	11
1.10. Causality in text	12
1.11. Coherence/Cohesion of conjunctions	15
1.12. Degrees of finiteness	16
1.13. Study of POS bigrams	17
1.14. Position of function words in sentence	19
1.15. Noun phrase	19
1.16. Length of R-motifs	21
1.17. Length of D-motifs	23
1.18. Syntactic complexity	25
1.19. Adnominal modifiers - 0. Classification	26
1.20. Adnominal modifiers - 1. Weight	28
1.21. Adnominal modifiers - 2. Distribution	29
1.22. Adnominal modifiers - 3. Complexity	31
1.23. Adnominal modifiers - 4. Cohesion	32
1.24. Adnominal modifiers - 5. Scaling	33
1.25. Adnominal modifiers - 6. Motifs	34
<b>2. Semantics</b>	<b>36</b>
2.1. Polysemy in German	36
2.2. Polysemy in the Swadesh list	37
2.3. The course of polysemy in sentence	38
2.4. Polysemic similarity and distance	39
2.5. Polysemic text construction	40
2.6. Vectors of polysemy	42
2.7. Synonymy	43
2.8. Semantic classes of nouns	44
2.9. Semantic diversification of prefixes	46

2.10. Semantics of prepositions 1	48
2.11. Semantics of prepositions 2	49
2.12. Degree of metaphoricality	50
2.13. Distribution of metaphoricality	52
2.14. Metaphoricality and frequency	53
2.15. Metaphoricality and length	54
2.16. The weight of metaphoricality in text	55
2.17. Metaphoricality motifs	56
2.18. Modality marking	57
2.19. Modality degree and frequency	59
2.20. Sound symbolism	60
2.21. Association of antonyms	62
2.22. Adjectival antonyms	64
2.23. B-motifs	66
<b>3. Textology</b>	<b>69</b>
3.1. Style	69
3.2. Style evolution	70
3.3. Entropy deployment	71
3.4. The type-token relation	73
3.5. Stage play analysis 1	75
3.6. Stage play analysis 2	76
3.7. Text properties	77
3.8. Thematic words and Frumkina's law	78
3.9. Distances in text	80
3.10. Text cohesion	83
3.11. Arc length of frequencies	85
3.12. Sentence sequences	86
3.13. Verbal antonymy	88
3.14. Hurst exponent	89
3.15. Hreb construction	92
<b>4. Pragmatics</b>	<b>94</b>
4.1. Speech act distribution	94
4.2. Speech act motifs	95
4.3. Speech act length	97
4.4. R-motifs of speech acts	98
4.5. Scaling of speech acts	100
<b>5. Synergetics</b>	<b>102</b>
5.1. Word length and polysemy	102

5.2. Kelih's Repeat Rate hypothesis	103
5.3. Word length and compositionality	104
5.4. Allomorphic complexity	105
5.5. Control cycle	106
<b>6. Various issues</b>	<b>108</b>
6.1. Scaling of dogmatism	108
6.2. Word frequency and initial clusters	109
6.3. Frequency and position in text	110
6.4. Hapax legomena and synthetism	110
6.5. Compound degree	112
6.6. Diversification theory	113
6.7. Givón's hypothesis	115
6.8. Laws in language	115
6.9. Morphological complexity of words	118
6.10. Syllable length	119
6.11. Nominal compounding tendencies in German	121
6.12. Quantification exercise	123
6.13. Corpus linguistics and theory	125
6.14. Small inventories	126
6.15. Borrowing	129
6.16. Frequency and irregular verbs	130
6.17. Ord's plane	132
6.18. Block distribution of modal expressions	132
6.19. Sonority sequences	135
6.20. Verb classes	137
<b>Author index</b>	<b>139</b>
<b>Subject index</b>	<b>146</b>