Problems
in
Quantitative Linguistics
4

by
Reinhard Köhler
Gabriel Altmann

2014
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)


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
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*, *Glotto-theory* 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**

**1. Syntax**

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 modificators - 0. Classification 26  
1.20. Adnominal modificators - 1. Weight 28  
1.22. Adnominal modificators - 3. Complexity 31  
1.23. Adnominal modificators - 4. Cohesion 32  
1.25. Adnominal modificators - 6. Motifs 34  

**2. Semantics**

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

6. Various issues 108

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

Author index 139
Subject index 146