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. Čech, G. Altmann, G., *Problems in quantitative linguistics 3.* 2011, VI + 168 pp.

ISBN: 978-3-942303-08-8

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 prolific 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 Fengxiang Fan and Eric Wheeler who read the book and helped us to improve our style. Radek Čech was supported by the Czech Science Foundation, grant no. P406/11/0268.

Radek Čech
Gabriel Altmann

Contents