Quantitative Linguistics, an Invitation

Karl-Heinz Best

Otto Rottmann

2017

RAM-Verlag

Studies in Quantitative Linguistics

Editors

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.
14. R. Köhler, G. Altmann, *Problems in Quantitative Linguistics Vol. 4.* 2014, VIII+148 pp.
15. Best, K.-H., Kelih, E. (eds.), *Entlehnungen und Fremdwörter: Quantitative Aspekte*. 2014. VI + 163 pp.
16. I.-I. Popescu, K.-H. Best, G. Altmann, G. *Unified Modeling of Length in Language*. 2014, VIII + 123 pp.
17. G. Altmann, R. Čech, J. Mačutek, L. Uhlířová (eds.), *Empirical Approaches to Text and Language Analysis dedicated to Luděk Hřebíček on the occasion of his 80th birthday*. 2014. VI + 231 pp.
18. M. Kubát, V. Matlach, R. Čech, *QUITA Quantitative Index Text Analyzer*. 2014, VII + 106 pp.

19. K.-H. Best, *Studien zur Geschichte der Quantitativen Linguistik.* 2015. III + 158 pp.
20. P. Zörnig, K. Stachowski, I.-I. Popescu, T. Mosavi Miangah, P. Mohanty, E. Kelih, R. Chen, G. Altmann, *Descriptiveness, Activity and Nominality in Formalized Text Sequences.* 2015. IV+120 pp.
21. G. Altmann, *Problems in Quantitative Linguistics Vol. 5*. 2015. III+146 pp.
22. P. Zörnig, K. Stachowski, I.-I. Popescu, T. Mosavi Miangah, R. Chen, G. Altmann, *Positional Occurrences in Texts: Weighted Consensus Strings.* 2015. II+178 pp.
23. E. Kelih, R. Knight, J. Mačutek, A.Wilson (eds.), *Issues in Quantitative Linguistics Vol. 4*. 2016. III + 231 pp.
24. J. Léon, S. Loiseau (eds.), *History of quantitative linguistics in France*. 2016. II + 232 pp.

ISBN: 978-3-942303-51-4

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

Preface

The present book is a strongly extended translation of the German original „Quantitative Linguistik. Eine Annäherung", 3rd edition. (Göttingen: Peust & Gutschmidt Verlag 2006). It is written for beginners in a rather "non-mathematical" language and contains many computations showing the procedures which should be followed by a researcher who begins to work with a language. It aims at showing some law-candidates concerning length, diversification, evolution, borrowings, ranking, mutual relations, first steps in synergetics, etc. A hypothesis can become law only if it is derived from a theory and corroborated on as many texts and languages as possible. An introduction to the history of quantitative linguistics shows that there is a very old tradition beginning about 2500 years ago.

Writing an introduction one never knows whether the contents and the way of presentation will attract or discourage the young readers. On the other hand, there are a small number of introductions which can show the reader some backgrounds of quantitative linguistics. Usually, the authors fulfill the books by mathematics, deriving of models, statistical tests, computing some probabilities etc. and the beginner does not know what to do. The present book shows how to define, how to count, how to fit a function or a distribution, how to evaluate it and how to present it in a publication. The rich bibliography attached to the book helps the reader to find ready evaluations, descriptions of the problem, the way of attaching the results to an existing theory, etc. The authors refer to many works written by their students and published in form of articles or dissertations.

The main motive of the book is: Begin and do not cease! Mathematics is no misfortune, on the contrary, it is a means to exactly express our findings and attach them to other ones. Even simple quantitative data collections are better than speaking about something in a non-formal language without any trials for corroboration.

The editorial board of this series recommends both to beginners and to advanced scholars to imitate the results using as many languages as possible in order to corroborate the existing results.

Gabriel Altmann

Contents

Preliminary Remarks

The following explications are addressed to those readers who are often “missed“ by authors and researchers whose specialty is Quantitative Linguistics (QL), but represent a large potential circle of readers and staff members: philologists (linguists, but also literary scholars and historians), who have never thought of seriously dealing with phenomena of linguostatistics, partly, since they did not allocate any relevance to such subjects, partly, since they assume that the methods required can only be learned elaborately and with uncertain results. In contrast to requirements in the tuition of physicians, psychologists, sociologists and economists the subjects of the philosophical faculties mostly do without the recommendation of acquiring knowledge in statistics or even the integration of the subject into the relevant courses of studies. Thus, a hurdle is developed which is an extensive obstacle for most representatives as to access to the possibilities of QL. The “success“ of that situation is that thematically versatile international research with an orientation to language statistics exists, which, however, is almost exclusively taken note of by specialists, not by the majority of philologists, though they could benefit from it. On the other hand, there exist evident needs with many linguists and literary scholars concerning precise statements related to their observations which find expression in statistics for most different objects, e.g. if attempts are made to describe stylistic particularities of a text, text type / genre or author.

Now, it has to be shown that and how it is possible to acquire part of the relevant research and even do own research work without long-time studies of statistics. That is made possible by the large improvement of the tools of the disciplines mentioned, especially if an interested scholar is ready to participate in cooperations. This means that it has to be attempted to remove some of the hermetic character from the quantitative studies in linguistics and literature. To achieve that the following subjects will be dealt with:

- development of quantitative linguistics (and studies in literature);
- subjects which can be processed quantitatively thus furnishing scientific progress;
- statistical observations concerning German vocabulary;
- detection of linguistic laws;
- theory;
- perspectives.

The above subjects are mainly described from the perspective of quantitative linguistics (QL) in the German speaking countries; despite that restriction the object can only be dealt with in excerpts and by means of examples. The main subject is the validation of law-like hypotheses. “Quantitative Literaturwissenschaft“[1] – a

[1] Quantitative studies in literature

term used by the German physicist Wilhelm Fucks (1968: 77, 88) – will play a marginal role only; mentioning it has a more programmatic character. However, it has to be stated that many questions dealt with in quantitative linguistics can be applied to literary texts and also to other genres of art without problems. "Quantitative literary science can teach us to understand better what actually takes action in the author or is done by him when writing his works by the mapping of issues in texts onto mathematical models. Generally, however, he is hardly aware of what he does formally."[2] (Fucks 1968: 88). This corresponds to all authors, not only the literary ones. A lot of information on what quantitative studies in literature can do is found in Altmann (1988a) and Altmann & Altmann (2005), Popescu et al. (2015).

Though explications are mainly addressed to non-specialized readers, and this is done in the hope that some interest can be aroused in them, specialists should profit from reading this book as well; they may confidently skip the trivial descriptions of elementary information. Dealing with linguistic laws concentrates on those whose testing the authors of this book were or still are somehow involved in the course of the projects performed in Germany, Austria and China. Many results appear there for the first time; with respect to subjects already published improved files have been elaborated or new calculations have been performed in some cases already. To obtain a depiction that is as demonstrative as possible almost all files are presented in the form of tables or graphical illustrations. It is not originality that is the primary target of these explications, but an overview that is as compact as possible; understandability is mainly intended. As contributions to quantitative linguistics can hardly be found in more widely circulating professional publications in linguistics to allow the further induction into subjects mentioned here, their theoretical backgrounds and the mathematical aspects in a lot of references are given to be used, if required.

Due to the restrictions concerning subjects it is not claimed that this book can be an introduction into quantitative linguistics being sufficient for all purposes. With respect to that important subjects like linguistic typology, stylistics and readability research come off badly. But all beginners can use it directly and extend the store of languages scrutinized up to now. Any analysis performed with the presented means would either corroborate a hypothesis or force us to revise it.

We'd like to recommend that the readers use the book *Quantitative Linguistics. An International Handbook*, edited by Köhler, Altmann and Piotrowski (2005), containing many domains of linguistics.

The present book is a strongly revised version of the third edition of K.-H. Best, *Quantitative Linguistik. Eine Annäherung*. Göttingen: Peust & Gutschmidt Verlag, 2006.

[2] „Eine quantitative Literaturwissenschaft kann uns durch die Abbildung von Sachverhalten in Texten auf mathematische Modelle besser verstehen lehren, was eigentlich im Autor vorgeht oder von ihm getan wird, wenn er seine Werke verfaßt. Dabei wird der Autor sich dessen, was er formal tut, im allgemeinen kaum bewußt sein.“