

Issues in Quantitative Linguistics
3

edited by

Reinhard Köhler
Gabriel Altmann

*Dedicated to Karl-Heinz Best
on the occasion of his 70th birthday*

2013
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.
13. R. Köhler, G. Altmann, *Issues in Quantitative Linguistics Vol 3*. 2013. II + 403 pp.

ISBN: 978-3-942303-12-5

© Copyright 2013 by RAM-Verlag, D-58515 Lüdenscheid
RAM-Verlag
Stüttinghauser Ringstr. 44
D-58515 Lüdenscheid
RAM-Verlag@t-online.de
<http://ram-verlag.de>

Karl-Heinz Best

was born on January 28, 1943 in Koblenz, Germany. He studied general linguistics and German linguistics at the University of Bonn (1964 - 1966), then in Bochum (1966 -1971) where he extended his study to Scandinavian languages. In Bochum he wrote his PhD thesis on analogy. From 1974 he taught at the University of Göttingen.

There are many important aspects of Best's scientific life. The first is the application of quantitative methods to writing, morphology, lexicon, syntax, language learning, language change, borrowing, language comprehension and history of quantitative linguistics. His main research field is German and Northern Germanic languages. His immortal merit is the testing of candidates for language laws on an enormous amount of data. The data were either collected by him from modern texts or found by him on places where nobody had ever sought them: in the history of German linguistics. In this way he has shown that German linguists were always interested in the quantitative evaluation of linguistic data. The 26 historiographic articles in his bibliography were published mostly in the journal *Glottometrics*, which enjoys his membership in the editorial board. Best's biographies bring to light the fact that linguistics was not that qualitative as presented in books about its history. There were ideas and works which were simply ignored because of the weight of other paradigms such as the neogrammarians and structuralism. The same was the fate of G.K. Zipf, whose philosophy passed almost unnoticed along the army of structural and generative linguists, who were sure to have discovered the "truth". Today, the traces of Zipf's work can be found in every scientific discipline, his hypotheses form the basis of frequentism and synergetic linguistics. Karl-Heinz Best detected the son of the famous physician S.V. Čebanov, who introduced the Poisson process into linguistics and wrote a historical note about his father together with him.

The second important activity of K.H. Best is the propagation of quantitative linguistics on the Internet. There are hundreds of articles written by him about different issues both in qualitative and quantitative linguistics. Not everybody is ready or takes the time to write popular articles in Wikipedia or Wiktionary but K.H. Best does it with relentless enthusiasm. In linguistic circles one says that the information about linguistics which can be obtained on the Internet is the "Best" one.

The third aspect is Best's editorial activity. This is not restricted to his membership in the editorial boards of *Glottology*, *Glottometrics* and *Göttinger Beiträge zur Sprachwissenschaft*, where he cannot impel the colleagues directly but he edits also omnibus volumes where he can play the role of a slave driver. His volumes are always ready on time and concern both synchronic and diachronic linguistics, everything nicely wrapped in a quantitative dress.

K.-H. Best cooperates at the conferences on *Knowledge transfer* in Göttingen and Halle.

Since we all begin to be active somewhere at the university and know how difficult it is to persuade colleagues and students that exact linguistics striving for theories is perhaps more advanced than concept coining, classifications and rule description, we cannot sufficiently emphasize the enormous success with which K.-H. Best won both students and colleagues for quantitative linguistics. Under normal circumstances one needs a Smith & Wesson, calibre 8 mm, to persuade a linguist that there are many ways to truth (not only generative linguistics) but mostly not even a revolver helps. But K.-H. Best succeeded without bodily violence and both his students and his colleagues published a number of articles (see References) which would fill several thick volumes. He called his collection *Göttinger Projekt Quantitative Linguistik* which can be found also on the Internet. Göttingen was one of the few German universities accepting quantitative analyses as bachelor, master or doctoral theses. His own work comprises more than 140 articles and more than 100 book reviews in the journal *Germanistik*.

Issues in Quantitative Linguistics Vol. 3
(= Studies in Quantitative Linguistics No. 13)
Contents

| | |
|--|-----|
| Karl-Heinz Best | 3 |
| Bibliography of K.-H. Best | 5 |
| Bibliography of the Göttingen Project (Pupils) | 16 |
| Gabriel Altmann | |
| Aspects of word length | 23 |
| Lu Wang | |
| Word length in Chinese | 39 |
| Emmerich Kelih | |
| Bemerkungen zur Wissenschaftsgeschichte der quantitativen Linguistik: Baudouin de Courtenay und Hugo Schuchardt | 54 |
| Peter Grzybek | |
| Homogeneity and heterogeneity within language(s) and text(s): Theory and practice of word length modelling | 66 |
| Kamil Stachowski | |
| The influx rate of Turkic glosses in Hungarian and Polish post-mediaeval texts | 100 |
| Peter Zörnig | |
| Distances between words of equal length in a text | 117 |
| Irma Sorvali | |
| Ecotranslatorology | 130 |
| Fan Fengxiang, Yu Yang, Wang Hua | |
| Subjectival position and syntactic complexity in English sentences | 137 |
| Relja Vulcanović | |
| Efficiency of Word Order in Flexible Parts-of-Speech Systems | 150 |
| Veronika Vincze | |
| The relationship of dependency relations and parts of speech in Hungarian | 168 |

| | |
|---|-----|
| Olexandra Nuzban, Sergej Kantemir | |
| Statistical analysis of perception adjectives ‘soft’ – ‘hard’ in English | 177 |
| Andrei Beliankou, Reinhard Köhler, Sven Naumann | |
| Distribution of the depth of argumentation relations | 195 |
| George K. Mikros | |
| Systematic stylometric differences in men and women authors: a corpus-based study | 206 |
| Ioan-Iovitz Popescu, Sven Naumann, Emmerich Kelih, Andrij Rovenchak, Anja Overbeck, Haruko Sanada, Reginald Smith, Radek Čech, Panchanan Mohanty, Andrew Wilson, Gabriel Altmann | |
| Word length: aspects and languages | 224 |
| Ján Mačutek, Gejza Wimmer | |
| Alternative methods of goodness-of-fit evaluation applied to word length data | 282 |
| Gordana Đuraš, Ernst Stadlober, Emmerich Kelih, Peter Grzybek | |
| Komplexität sprachlicher Formen: Die Singh-Poisson-Verteilung: ein Modell in der Wortlängenforschung? | 291 |
| Bahar Karaođlan, Bekir Taner Dinçer, Tarık Kışla, Senem Kumova Metin | |
| Language and corpus independent performance metric for language models | 309 |
| Sigurd Wichter | |
| Wie kommuniziert eine Gesellschaft aus linguistischer Sicht? | 326 |
| Petra Claudia Steiner | |
| Diversification of English valency patterns | 369 |
| Tim Duwaerts, Gereon Ullmann | |
| Quantitative Untersuchungen zur Valenz deutscher Substantive | 392 |

Quantitative linguistics (QL) is a sub-discipline of general linguistics and, more specifically, of mathematical linguistics. Quantitative linguistics deals with language learning, language change, and application as well as structure of natural languages. Finally, all these definitions form a hierarchy of more and more general meanings, whereby the number of definitions decreases with increasing generality. Among the levels of this kind of hierarchy, there exists a number of lawful relations. Menzerath's law (also, in particular in linguistics, Menzerath-Altmann law): This law states that the sizes of the constituents of a construction decrease with increasing size of the construction under study. Issues in Quantitative Linguistics 3. edited by Reinhard Köhler Gabriel Altmann. Dedicated to Karl-Heinz Best on the occasion of his 70. 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. Studies in Quantitative Linguistics. 6. Arjuna Tuzzi Ioan-Iovitz Popescu. Gabriel Altmann. Quantitative Analysis of. Italian Texts. RAM - Verlag. Quantitative Analysis of. Italian Texts. by Arjuna Tuzzi Ioan-Iovitz Popescu Gabriel Altmann. 2010 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) Jörn Mačutek (jmacutek@yahoo.com) Eric S. Wheeler (wheeler@ericwheeler.ca). 5. R. Köhler (ed.), Issues in quantitative linguistics. 2009, VI + 205 pp. 6. A. Tuzzi, I.-I. Popescu, G. Altmann, Quantitative analysis of Italian texts. 2010, IV+161 pp. © Copyright 2010 by RAM-Verlag, D-58515 Ländenscheid. Issues in Quantitative Linguistics 3. edited by Reinhard Köhler Gabriel Altmann. Dedicated to Karl-Heinz Best on the occasion of his 70. By providing Hungarian and Polish data, it also adds to the issue of Turkic influence in Europe, first discussed quantitatively in Best (2005) using the example of German. On the other hand, it attempts to show how the quantitative and qualitative DQDOVHV FRPSOHPHQW UDWKHU WKDQ RSSRVH RU H[FOXGH HDFK RWKHU /XGÄCEN +ÄœHEtÄ¾HN worked to demonstrate this to a more traditionalistic audience (1990: 371). My aim here is to illustrate how the quantitative approach can reveal a general tendency in a collection of detailed observations gathered and explained with the philological method. While the formal branches of linguistics use only the qualitative mathematical means (algebra, set theory) and logics to model structural properties of language, quantitative linguistics (QL) studies the multitude of quantitative properties which are essential for the description and understanding of the development and the functioning of linguistic systems and their components. The objects of QL research do, therefore, not differ from those of other linguistic and textological disciplines; nor is...