

Unified Modeling of Length in Language

Ioan-Iovitz Popescu

Karl-Heinz Best

Gabriel Altmann

2014

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 (eds.), *Issues in Quantitative Linguistics Vol 3*. 2013, IV + 403 pp.
14. R. Köhler, G. Altmann, *Problems in Quantitative Linguistics 4*. 2014, VI + 148 pp.
15. K.-H. Best, E. Kelih (eds.), *Entlehnungen und Fremdwörter: Quantitative Aspekte*. 2014, IV + 163 pp.
16. I.-I. Popescu, K.-H. Best, G. Altmann, *Unified modeling of length in language*. 2014, II + 124 pp.

ISBN: 978-3-942303-26-2

© 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 aim of this volume is to find a unified model of length distribution of any unit in language. It is merely a trial leaning against published results. It is impossible to perform the analysis for “all possible” units because only some of them are known and used, the definition of new ones is the normal policy in science. And even if a new unit is defined, it takes a special place in the hierarchy of other units which must be defined, too. Hence, this enterprise is endless.

We used published data from ca 50 languages, their dialects and historical epochs. It was impossible to use everything that has ever been published, it is already a separate discipline within linguistics. The majority of data has been elaborated in several projects performed in Göttingen where also students took part in the work (cf. <http://www.gwdg.de/~kbest/projekt.htm>). Besides three volumes quoted in the references (cf. Best 1997, 2001; Grzybek 2006) there are a number of publications in different journals.

Our main aim was to find a law of length and its special forms at individual language levels, that is, to avoid the search for ever new models whose validity is always merely local: they hold for the given level in the given language and taken together form an enormous family of distributions. In the unified model there are merely differences in the parameters, and the parameters themselves are part of a dynamic system displaying self-regulation.

We hope that researchers will test it in further languages and on different linguistic levels.

Ioan-Iovitz Popescu
Karl-Heinz Best
Gabriel Altmann

Contents

1. Introduction	1
2. Syllable length	7
3. Morph length	11
4. Word length	14
5. Length of compounds	87
6. Length of rhythmic units	89
7. Verse length in words	91
8. Sentence length	94
9. Speech act chains	108
10. The levels	110
11. Conclusion	111
Appendix	112
References	114
Author index	120
Subject index	122

