

Descriptiveness, Activity and Nominality in Formalized Text Sequences

Peter Zörnig

in cooperation with

**Kamil Stachowski
Ioan-Iovitz Popescu
Tayebah Mosavi Miangah
Panchanan Mohanty
Emmerich Kelih
Ruina Chen
Gabriel Altmann**

**2015
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 (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. Mčutek, 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
19. K.-H. Best, *Studien zur Geschichte der Quantitativen Linguistik*. 2015. III + 158 pp.

ISBN: 978-3-942303-31-6

© Copyright 2015 by RAM-Verlag, D-58515 Lüdenscheid

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

Preface

In the present book we study characteristics of language based on formalized text sequences. The study of text as a sequence of various entities is rapidly developing in form of articles, omnibus volumes and monographs. In fact, our linguistic study can be considered as a part of a very fertile interdisciplinary research activity devoted to the analysis of information sequences. Such sequences occur also in computational biology (e.g. in form of DNA strings), in coding theory and data compression. While qualitative linguistic analysis searches for rules which are important for language learning, quantitative analysis tries to capture hidden mechanisms which are not necessary for the understanding of language. Except for certain poetic phenomena, e.g. rhythm which can be produced consciously, these mechanisms cannot be learned and do not represent the core of standard linguistics.

In the present book, a group consisting of mathematicians and linguists – specialists for a certain language – attempts to discover textual phenomena which may seem to be strange for the “normal” linguistics but whose deciphering may help to reveal candidates for laws. Laws are the highest aim of science because without them no theories and no explanations are possible. Unfortunately, in linguistics the testing of a hypothesis is never finished, one can at most validate it to a certain degree. In practice, this validation will never terminate because one would be forced to analyze all languages and, in case of text laws, as many texts as possible. Here no corpuses can help because none of them contains the complete history of language, the evolution of an individual speaker or a complete collection of text sorts.

Hence our attempts merely reveal a few of the infinite number of facets of a text. We try to collect data, find models of their behavior in form of hypotheses, test them, compare the results in texts of eleven languages available to us and try to create a research domain which will never be satisfactorily explored.

We present all observed data in order to enable other researchers to analyze them applying other methods or other characterizations, and to formulate and test other hypotheses. We reduced the whole field to specific phenomena of description, activity and specifying, otherwise the study would be too extensive. Nevertheless, we show at some places the possibility of going into the depth of the hierarchy of phenomena.

Peter Zörnig

Contents

1.	Introduction	1
2.	Descriptiveness vs. activity	4
	2.1. Definitions and tests	4
	2.2. Sequential measurement	19
	2.3. Runs of two elements	33
3.	Nominality	40
	3.1. Nominality vs. predicativity	40
	3.2. Variability	45
	3.3. Triads	46
	3.4. Runs of three elements	49
4.	Distances	53
5.	Some further aspects	72
	5.1. Predicativity motifs	72
	5.2. Length and frequency	74
	5.3. Rank-frequency of predication motifs	83
6.	Conclusions	97
	References	99
	Appendix	102
	Author index	118
	Subject index	119