Studies
in
Quantitative Linguistics
8

Ioan-Iovitz Popescu
Ján Mačutek
Emmerich Kelih
Radek Čech
Karl-Heinz Best
Gabriel Altmann

Vectors and Codes of Text

RAM - Verlag

Vectors and codes of text

by

Ioan-Iovitz Popescu

in cooperation with

Ján Mačutek
Emmerich Kelih
Radek Čech
Karl-Heinz Best
Gabriel Altmann

2010
RAM-Verlag

Studies in quantitative linguistics

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, II + 161 pp.

Gedruckt mit Unterstützung der Karl-Franzens-Universität Graz

ISBN: 978-3-942303-02-6

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

Preface

The present book is a continuation of our endeavour to introduce in textology new quantitative methods and evaluate some older ones (cf. Popescu et. al. 2009, Popescu, Mačutek, Altmann 2009; Tuzzi, Popescu, Altmann 2010). We illustrated the measurements and performed evaluations of texts from many languages. Needless to say, all results ensuing from the use of vectors, codes and chains must be tested on further languages and texts. Since this is ongoing empirical research, some modifications and adaptations of the methods presented may be necessary.

Nevertheless, the more we advanced the clearer we saw the abysmal playground hidden in texts. With some progress in using and elaborating quantitative methods in text analysis some new problems and formerly unrecognized phenomena appeared, thus we were confronted not only with methodological challenges, but also with new questions and problems in linguistic text theory in general.

We restricted ourselves to formal features (frequencies, codes and chains) accessible to the cooperating linguists and avoided sociolinguistic, psycholinguistic and other problems. We nevertheless hope that the methods presented could be made useful for other investigations, too.

The book consists of nine chapters. In Chapter 1 we introduce briefly the extensive domain of possible problems concerning comparisons and research strategies. In Chapters 2 to 6 we examine different vectors of texts, show their behaviour, compare texts and languages and take a step towards capturing the text dynamics looking at it from different points of view.

In Chapters 7 and 8 we goedelize the text in one special way, show breaks in the syntactic continuity in the text and ascribe to it its binary code which can be compared and tested.

The last chapter, Chapter 9, is devoted to chaining phenomena restricted here to Belza-Skorochoďko chains, revealing many new vistas touching behaviour, perseveration, psycholinguistic and other aspects. As a matter of fact, each topic could be developed infinitely but we strived for presenting simple methods, developed tests and showed a way of ternary plotting.

We hope that other scholars will adopt the methods for different purposes and for analyzing other languages, in order to get stronger corroboration of the procedures presented.

In this place we want to express our gratitude to Claudiu Vasilescu, who patiently wrote for us dilettantes all the Excel programmes and discretely concealed his amazement about our naivety. We had suffered much more without his kind help. Radek Čech was supported by the Czech Science Foundation, grant no. 405/08/P157 – Components of transitivity analysis of Czech sentences (emergent grammar approach).

I.-I. Popescu

Contents