

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)
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.

ISBN: 978-3-942303-00-2

© Copyright 2010 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 present book can be considered a continuation of two streams in textology initiated by different groups of scholars. The first group consisting of Italian scientists analyzed the 57 end-of-year speeches of 9 Italian presidents since 1949 to 2005 from very different points of view. The investigation resulted in the book edited by M.A. Cortelazzo and A. Tuzzi (2007) containing some statistical vistas but concentrated mostly to political, social, linguistic and cultural matter. The second group was lead by I.-I. Popescu and consisted of linguists and mathematicians who concentrated rather to general aspects associated with texts, namely text characterization, fixed points, vocabulary richness, thematic concentration, compactness, study of autosemantics, the impact of the position in sentence, but also to the possibility of drawing conclusions from text to the morphology of language, etc. This investigation resulted in two books authored by I.-I. Popescu et al. (2009) and I.-I. Popescu, J. Mačutek and G. Altmann (2009).

The present book is a further development of methods introduced by the second stream but fully concentrated to the end-of-year speeches of Italian presidents. Moreover, the updated version of the corpus includes 60 end-of-year speeches of 10 presidents since 1949 to 2008. Our aim was to look at the highest floor of the linguistically perceived reality, namely the sets of word associations, compare these worlds and pursue a possible development in this thematically homogeneous complete corpus. Of course, not everything could be made but besides associations we added some chapters concerning the golden section, Zipf's law, parts-of-speech analysis and some issues concerning vocabulary richness, continuing the examinations resulting from earlier common publications.

Acknowledgement

We want to express our gratitude especially to Mrs. Mihaela Vasilescu for her kind help in computing affairs in places where our abilities seemed to be exhausted. She followed our motto: *viribus unitis*.

A.T., I.-I.P., G.A.

Contents

Preface	I
1. Introduction	1
2. Associative analysis	5
2.1. Introduction	5
2.2. Associative concentration	7
2.3. The indicator C_{rel}	11
2.4. The indicator κ_{rel}	13
2.5. The levels	16
2.6. Cliques	19
2.7. The associated vocabulary	36
2.7.1. Thematic concentration	36
2.7.2. Thematic association	41
2.8. Degree distribution	44
2.8.1. Degree spectrum	44
2.9. The comparison of vocabularies	51
2.9.1. Method I	51
2.9.2. Method II	82
2.9.3. Conclusion	94
3. The golden section	95
4. Zipf's law	107
4.1. Validity	107
4.2. Stratificational approach	112
5. Parts-of-speech diversification	116
5.1. Zipf's law	116
5.2. The indicators p and q	123
6. Vocabulary richness	127
7. Vector comparisons	135
References	155
Author index	158
Subject index	160