Qualico 2009
Text and Language

Structures · Functions · Interrelations
Quantitative Perspectives

Edited by
Peter Grzybek
Emmerich Kelih
Ján Mačutek

Advisory Editor
Eric S. Wheeler

Praesens Verlag
Wien 2010
Contents

Preface vii
Peter Grzybek, Emmerich Kelih, Ján Mačutek

Quantitative analysis of Keats’ style: genre differences 1
Sergej Andreev

Word-length-related parameters of text genres in the Ukrainian language. 13
Solomiya Buk, Olha Humenchyk, Lilija Mal’seva, Andrij Rovenchak

On the quantitative analysis of verb valency in Czech 21
Radek Čech, Ján Mačutek

A link between the number of set phrases in a text and the number of described facts 31
Łukasz Dębowski

Modeling word length frequencies by the Singh-Poisson distribution 37
Gordana Đuraš, Ernst Stadlober

How do I know if I am right? Checking quantitative hypotheses 49
Sheila Embleton, Dorin Urîtescu, Eric S. Wheeler

Text difficulty and the Arens-Altmann law 57
Peter Grzybek

Parameter interpretation of the Menzerath law: evidence from Serbian 71
Emmerich Kelih

A syntagmatic approach to automatic text classification. Statistical properties of $F$- and $L$-motifs as text characteristics 81
Reinhard Köhler, Sven Naumann

Probabilistic reading of Zipf 91
Jan Králik

Revisiting Tertullian’s authorship of the Passio Perpetuæ through quantitative analysis 99
Jerónimo Leal, Giulio Maspero

Textual typology and interactions between axes of variation 109
Sylvain Loiseau
vi Contents

Rank-frequency distributions: a pitfall to be avoided
Ján Mačutek 119

Measuring lexical richness and its harmony
Gregory Martynenko 125

Measuring semantic relevance of words in synsets
Ivan Obradović, Cvetana Krstev, Duško Vitas 133

Distribution of canonical syllable types in Serbian
Ivan Obradović, Aljoša Obuljen, Duško Vitas, Cvetana Krstev, Vanja Radulović 145

Statistical reduction of the feature space of text styles
Vasilij V. Poddubnyj, Anastasija S. Kravcova 159

Quantitative properties of the Nko writing system
Andrij Rovenchak, Valentin Vydrin 171

Distribution of motifs in Japanese texts
Haruko Sanada 183

Quantitative data processing in the ORD speech corpus of Russian everyday communication
Tatiana Sherstinova 195

Complex investigation of texts with the system “StyleAnalyzer”
O.G. Shevelyov, V.V. Poddubnyj 207

Retrieving collocational information from Japanese corpora: its methods and the notion of “circumcollocate”
Tadaharu Tanomura 213

Diachrony of noun-phrases in specialized corpora
Nicolas Turenne 223

Subject index 237

Author index 243

Authors’ addresses 247