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

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

Word-length-related parameters of text genres in the Ukrainian language (a pilot study)
Solomija Buk, Olha Humenchyk, Lilija Mal’tseva, Andrij Rovenchak 13

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

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

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

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

Text difficulty and the Arens-Altman law
Peter Grzybek 57

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

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

Probabilistic reading of Zipf
Jan Králík 91

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

Textual typology and interactions between axes of variation
Sylvain Loisseau 109

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

Measuring lexical richness and its harmony 125
Gregory Martynenko

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

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

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

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

Distribution of motifs in Japanese texts 183
Haruko Sanada

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

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

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

Diachrony of noun-phrases in specialized corpora 223
Nicolas Turenne

Subject index 237
Author index 241
Authors’ Addresses 245