

Contributions to the Science of Text and Language

Word Length Studies and Related Issues

Edited by

Peter Grzybek

University of Graz, Austria

 Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-4069-5 (PB)
ISBN-13 978-1-4020-4069-6 (PB)
ISBN-10 1-4020-4067-9 (HB)
ISBN-13 978-1-4020-4067-2 (HB)
ISBN-10 1-4020-4068-7 (e-book)
ISBN-13 978-1-4020-4068-9 (e-book)77

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springeronline.com

Printed on acid-free paper

All Rights Reserved

© 2007 Springer

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed in the Netherlands

Contents

Preface	ix
1	
Introductory Remarks: On the Science of Language in Light of the Language of Science <i>Peter Grzybek</i>	1
2	
History and Methodology of Word Length Studies <i>Peter Grzybek</i>	15
3	
Information Content of Words in Texts <i>Simone Andersen, Gabriel Altmann</i>	91
4	
Zero-Syllable Words in Determining Word Length <i>Gordana Antić, Emmerich Kelih, Peter Grzybek</i>	117
5	
Within-Sentence Distribution and Retention of Content Words and Function Words <i>August Fenk, Gertraud Fenk-Oczlon</i>	157
6	
On Text Corpora, Word Lengths, and Word Frequencies in Slovenian <i>Primož Jakopin</i>	171
7	
Text Corpus as an Abstract Data Structure <i>Reinhard Köhler</i>	187
8	
About Word Length Distribution <i>Victor V. Kromer</i>	199
9	
The Fall of the Jers in the Light of Menzerath's Law <i>Werner Lehfeldt</i>	211

viii	<i>CONTRIBUTIONS TO THE SCIENCE OF TEXT AND LANGUAGE</i>	
10	Towards the Foundations of Menzerath's Law <i>Anatolij A. Polikarpov</i>	215
11	Aspects of the Typology of Slavic Languages <i>Otto A. Rottmann</i>	241
12	Multivariate Statistical Methods in Quantitative Text Analyses <i>Ernst Stadlober, Mario Djuzelic</i>	259
13	Word Length and Word Frequency <i>Udo Strauss, Peter Grzybek, Gabriel Altmann</i>	277
14	Developing the Croatian National Corpus and Beyond <i>Marko Tadić</i>	295
15	About Word Length Counting in Serbian <i>Duško Vitas, Gordana Pavlović-Lažetić, Cvetana Krstev</i>	301
16	Word-Length Distribution in Present-Day Lower Sorbian Newspaper Texts <i>Andrew Wilson</i>	319
17	Towards a Unified Derivation of Some Linguistic Laws <i>Gejza Wimmer, Gabriel Altmann</i>	329
	Contributing Authors	339
	Author Index	343
	Subject Index	347

Text, Speech and Language Technology

VOLUME 31

Series Editors

Nancy Ide, *Vassar College, New York*

Jean Véronis, *Université de Provence and CNRS, France*

Editorial Board

Harald Baayen, *Max Planck Institute for Psycholinguistics, The Netherlands*

Kenneth W. Church, *AT & T Bell Labs, New Jersey, USA*

Judith Klavans, *Columbia University, New York, USA*

David T. Barnard, *University of Regina, Canada*

Dan Tufis, *Romanian Academy of Sciences, Romania*

Joaquim Llisterri, *Universitat Autònoma de Barcelona, Spain*

Stig Johansson, *University of Oslo, Norway*

Joseph Mariani, *LIMSI-CNRS, France*