

Glottometrics 45

2019

RAM-Verlag

ISSN 1617-8351
e-ISSN 2625-8226

Glottometrics

Indexed in ESCI by Clarivate Analytics and SCOPUS by Elsevier

Glottometrics ist eine unregelmäßig erscheinende Zeitschrift (2-3 Ausgaben pro Jahr) für die quantitative Erforschung von Sprache und Text.

Beiträge in Deutsch oder Englisch sollten an einen der Herausgeber in einem gängigen Textverarbeitungssystem (vorrangig WORD) geschickt werden.

Glottometrics kann aus dem **Internet** heruntergeladen werden (**Open Access**), auf **CD-ROM** (PDF-Format) oder als **Druck-version** bestellt werden.

Glottometrics is a scientific journal for the quantitative research on language and text published at irregular intervals (2-3 times a year).

Contributions in English or German written with a common text processing system (preferably WORD) should be sent to one of the editors.

Glottometrics can be downloaded from the **Internet (Open Access)**, obtained on **CD-ROM** (as PDF-file) or in form of **printed copies**.

Herausgeber – Editors

G. Altmann	Univ. Bochum (Germany)	ram-verlag@t-online.de
S. Andreev	Univ. Smolensk (Russia)	smol.an@mail.ru
K.-H. Best	Univ. Göttingen (Germany)	kbest@gwdg.de
R. Čech	Univ. Ostrava (Czech Republic)	cechradek@gmail.com
E. Kelih	Univ. Vienna (Austria)	emmerich.kelih@univie.ac.at
R. Köhler	Univ. Trier (Germany)	koehler@uni-trier.de
H. Liu	Univ. Zhejiang (China)	lhtzju@gmail.com
J. Mačutek	Univ. Bratislava (Slovakia)	jmacutek@yahoo.com
A. Mehler	Univ. Frankfurt (Germany)	amehler@em.uni-frankfurt.de
M. Místecký	Univ. Ostrava (Czech Republic)	MMistecky@seznam.cz
G. Wimmer	Univ. Bratislava (Slovakia)	wimmer@mat.savba.sk
P. Zörnig	Univ. Brasilia (Brasilia)	peter@unb.br

External Academic Peers for Glottometrics

Prof. Dr. Haruko Sanada

Rissho University, Tokyo, Japan (<http://www.ris.ac.jp/en/>);

Link to Prof. Dr. Sanada: <http://researchmap.jp/read0128740/?lang=english>; <mailto:hsanada@ris.ac.jp>

Prof. Dr. Thorsten Roelcke

TU Berlin, Berlin, Germany (<http://www.tu-berlin.de/>)

Link to Prof. Dr. Roelcke: [https://www.daf.tu-](https://www.daf.tu-berlin.de/menue/deutsch_als_fremd_und_fachsprache/mitarbeiter/professoren_und_pds/prof_dr_thorst)

[berlin.de/menue/deutsch_als_fremd_und_fachsprache/mitarbeiter/professoren_und_pds/prof_dr_thorst_en_roelcke](https://www.daf.tu-berlin.de/menue/deutsch_als_fremd_und_fachsprache/mitarbeiter/professoren_und_pds/prof_dr_thorst_en_roelcke)

[mailto:Thosten.Roelcke \(roelcke@tu-berlin.de\)](mailto:Thosten.Roelcke@tu-berlin.de)

Bestellungen der CD-ROM oder der gedruckten Form sind zu richten an

Orders for CD-ROM or printed copies to RAM-Verlag RAM-Verlag@t-online.de

Herunterladen/ Downloading: <https://www.ram-verlag.eu/journals-e-journals/glottometrics/>

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Glottometrics. 45 (2019), Lüdenscheid: RAM-Verlag, 2019. Erscheint unregelmäßig.

Diese elektronische Ressource ist im Internet (Open Access) unter der Adresse

<https://www.ram-verlag.eu/journals-e-journals/glottometrics/> verfügbar.

Bibliographische Deskription nach 45 (2019)

online/ e-version ISSN 2625-8226 (print version ISSN 1617-8351)

Glottometrics 45, 2019

Abstracts

Emmerich Kelih

Segmental and Suprasegmental Vowel Frequencies in Slovene: Statistical Modeling 1

Abstract. We show that in Slovene, length, accent, and shortening of vowels represent factors influencing the frequency of vowels both in the dictionary and in texts. The results of the operation of these forces are presented by means of continuous models which are fitted to the resulting numbers.

Jan Hůla, Miroslav Kubát, Radek Čech, Xinying Chen, David Číž, Kateřina Pelegrinová, Jiří Milička

Context Specificity of Lemma. Diachronic Analysis

Abstract. This study deals with the recently proposed concept of so-called Context Specificity of Lemma (*CSL*). *CSL* is based on the word embedding technique called Word2vec which enables measuring lexical context similarity between lemmas. Specifically, a recently proposed method Closest Context Specificity (*CCS*) is applied to a diachronic analysis of Czech texts. This method expresses how unique is a context within which a given lemma appears. The aim of the paper is to study what kind of semantic features can *CCS* detect and how useful could *CCS* be in a diachronic semantic analysis. The second goal is to observe the relation of *CCS* to frequencies in the corpora.

Renkui Hou, Chu-Ren Huang, Mi Zhou, Menghan Jiang

Distance between Chinese Registers Based on the Menzerath-Altmann Law and Regression Analysis

Abstract. This paper proposes an innovative method/index to represent the formality of a register based on the Menzerath–Altmann law and regression analysis. This index also can be used to quantify the distance between two registers. Analysis demonstrates that average word length decreases with the increase of clause length in each register and that their relationship can be fitted by the formula $y = ax^b$. It can be shown that the link between average word length and clause length abides by the Menzerath–Altmann law. Texts were represented by the fitted parameters, a and b , and their positions were plotted in 2-dimensions. Linear regression can be used to fit the functional correlation between these two parameters in each register. We show that the a -intercept of this regression line can be used as an index to represent the formality degree of the register and to compute the distance between two registers.

Hanna Gnatchuk

The classification of English styles on the basis of lexical parameters:
a case of clustering analysis

Abstract: The present article is an attempt to reveal the groups of the most similar and dissimilar English styles (or genres) on the basis of three factors (variables): their average word repeat, hapax legomenas and the number of unique words. We intend here to perform a clustering analysis, which is grounded on the Euclidean distance matrix. In this research we have determined the number of clusters (= the groups) in which English styles can be divided. The results have been explained, considering Elbowplot and Dendrogram. The necessary calculations have been done in Programs R-Studio and Python.

Zheyuan Dai, Haitao Liu

Quantitative Analysis of Queen Elizabeth II and American Presidents' Christmas Messages Over 50 Years (1967-2018)

Abstract. Over the past century, UK and US have evolved new Christmas traditions, namely Queen's Christmas Broadcasts for UK and lighting the National Christmas Tree for US. Queen Elizabeth II and American Presidents deliver their Christmas felicitations accompanying with new celebrations respectively. This study intends to evaluate stylistic features, especially at lexical level, of Queen Elizabeth II and American Presidents' Christmas messages based on the material over 50 years both synchronically and diachronically. Results exhibit that Queen Elizabeth II overall has a higher level of vocabulary richness along the half century. Detailed indicators, big words and hapax legomena, further show that Queen Elizabeth II's vocabulary is more complex and diversified. Nevertheless, American Presidents surpass Queen Elizabeth II in thematic concentration. Discourse analysis discovers that Queen Elizabeth II concentrates on many smaller-scale themes without political ones and cares for accuracy of words. While in addition to conveying good wishes, American Presidents take Christmas messages as a good opportunity to publicize political opinions, leading to an overall higher thematic concentration level.

Yawen Wang, Haitao Liu

The Effects of Source Languages on Syntactic Structures
of Target Languages in the Simultaneous Interpretation:
A Quantitative Investigation Based on Dependency Syntactic Treebanks

Abstract. Dependency distance (DD), as the distance between two linked words in one sentence is widely used to explore the cognitive demands and cross-linguistic syntactic features in language processing. The purpose of simultaneous interpreting is to enable smooth communication between two languages, though it imposes a large burden on interpreters. However, previous studies have not yet investigated the impact of source languages on the target language in the simultaneous interpreting process between different language pairs from a typological perspective quantitatively. It is still indispensable to examine carefully how essential the role is played by different source languages in simultaneous interpreting. With recourse to quantitative methods, the current study explores the English simultaneous interpretations from distinct source languages. From the cognitive perspective, results via mean dependency distance demonstrate that the structures of English interpretations are

interfered marginally significantly by diverse source languages in simultaneous interpreting. Meanwhile, language typology of the source languages has moderately small impact on the English interpretations with resort to dependency direction. The current research firstly investigates the effect of diverse source languages on the same target language in simultaneous interpreting, suggesting the overwhelming impact of mean dependency distance minimization on language processing.

**Marija Radojičić, Biljana Lazić, Sebastijan Kaplar, Ranka Stanković,
Ivan Obradović, Ján Mačutek, Lívia Leššová**

Frequency and length of syllables in Serbian

Abstract. Basic analyses of several properties of syllables (the rank-frequency distribution, the distribution of length, and the relation between length and frequency) in Serbian is presented. The syllabification algorithm used combines the maximum onset principle and the sonority hierarchy. Results indicate that syllables behave similarly to words as far as mathematical models are concerned, but values of parameters in models for syllables are quite different from those for words.