

# **Glottometrics 46**

## **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

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

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

Die Deutsche Bibliothek – CIP-Einheitsaufnahme  
Glottometrics. 46 (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 46 (2019)  
**online/ e-version ISSN 2625-8226 (print version ISSN 1617-8351)**

# Glottometrics 46, 2019

## Abstracts

**Solomija Buk, Andrij Rovenchak**

**Simple definition of distances between texts from rank–frequency distributions.  
A case of Ukrainian long prose works by Ivan Franko**

**Abstract.** We present the analysis of long prose texts using several simple definitions of distance based on rank–frequency distributions. Various types of the Euclidean distance, the Jaccard distance, and the cosine distance are calculated. Our approach is useful for studies of groups of texts, where different definitions of distance show different relations with respect to the shortest and the longest text.

**Andrei V. Zenkov, Michal Místecký**

**The Romantic Clash: Influence of Karel Sabina over Mácha’s *Cikáni* from the  
Perspective of the Numerals Usage Statistics**

**Abstract.** The paper deals with the long-standing stylometric problem of Czech fiction – the authorship of the novel *Cikáni*. Although the text has been usually attributed to K. H. Mácha, there is a widespread hypothesis that its final shape was substantially influenced by a friend of his, K. Sabina. To solve the problem, we have exposed the works by Mácha (*Cikáni*) and Sabina (*Hrobník* and *Oživené hroby*) to the novel statistical attribution method, which takes into account the usage of numerals in texts. To provide a contrast to the new procedure, we have also employed a more conventional MFW analysis. The results, which are rather contradictory, are accounted for by various interpretations. The goal of the article is to show the soundness of the new method and check its applicability on Czech pieces of literature.

**Sergey Andreev**

**Types of Syllable Distribution in Russian Long Poems**

**Abstract.** In the present examination, syllable distribution in 15 long poems written during the period of the end of the 20<sup>th</sup> and the beginning of the 21<sup>st</sup> centuries by Russian authors is studied. The distribution of different types of syllables as well as the relationship of the initial and the last syllabic positions in verse lines are explored using one of non-parametric methods – the Kendall’s rank correlation coefficient. The syllabic types are formed by vowel–consonant sequences according to the principles of sonorant theory. The results revealed that the rank-frequency distribution of syllabic types in all poems under study is well fitted by the exponential function. Within the poems, the initial and final syllables were found to form a strong opposition, laying the basis for syllabic asymmetry of the verse line.

**Anna Rácová, Peter Zörnig, Gabriel Altmann**

### **Syllable structure in Romani**

**Abstract.** We present some methods to analyze tendencies that can be discovered in the syllable structure. To this end we study regularities in the Romani language as spoken in Slovakia. The results may be useful to classify languages and to support a future theory.

**Huiying Cai, Yunhua Qu, Zhiwei Feng**

### **A corpus-based study of the semantic prosody of Chinese light verb pattern across registers: taking *jinxing* and *shoudao* as examples**

**Abstract.** Light verbs constitute a fundamental element of the Chinese language. The construction “V+VN” was identified as a major light verb pattern (LVP) due to its high frequency of occurrence. Although the syntactic aspect of Chinese light verbs has been discussed in detail, the pragmatic aspect remains undetermined. Thus, this study examined *jinxing* ‘be in progress’ and *shoudao* ‘come in for’ as representatives of the two types of Chinese light verbs, agentive-action and accusative-action, respectively. Their patterns were regarded as complete entities of form and meaning. Register and semantic prosody, possessing an evaluative and attitudinal function, were introduced in order to explore the pragmatic aspect of Chinese light verbs. Based on a self-built corpus, the preference of semantic prosody created by the two typical LVPs and the influence of registers (spoken vs. written) on both, as well as on the semantic prosody created, were discussed in a quantitative manner. Furthermore, through the two general corpus approaches, variationist and text-linguistic, similarities and differences between the two LVPs themselves and their semantic prosody influenced by the registers were identified. The reasons for the corresponding results were discussed based on a situational analysis within the register analysis framework. The results broaden our understanding of Chinese light verbs from the syntactic aspect to the pragmatic aspect. Our findings would be substantially useful for Chinese language learning and teaching about the usage of Chinese LVPs in second language acquisition.

**Karl-Heinz Best, Michal Místecký, Peter Zörnig, Gabriel Altmann**

### **Quantifying the Quantitative Meter: On Rhythmic Types in the Dactylic Hexameter**

**Abstract.** In the present article, we show that the rank-order of the hexameter types used in the poetry productions in four languages (Greek, Latin, German, and Czech) abides by the exponential function. Up to now, more complex probability distributions have been used (cf. Best 2008, 2009), namely the negative hypergeometric and the Pólya ones, having one parameter more. We also establish the relation between the parameters of a given model.

**Michal Místecký, Gabriel Altmann**

**Tense and Person in English**

**Abstract.** In the article, ready-made data are used to find a simple model. The verbs are ordered in semantic classes, and for the ranked frequencies of tenses and verbal persons, some models are found. It is recommended to use as few parameters as possible. Here, the exponential and the Zipf-Alekseev functions are used, in one case the power function, too.