

# **Glottometrics 25**

## **2013**

**RAM-Verlag**

**ISSN 2625-8226**

# Glottometrics

**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 **Druckversion** 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>K.-H. Best</b>	Univ. Göttingen (Germany)	kbest@gwdg.de
<b>G. Djuraš</b>	Joanneum (Austria)	Gordana.Djuras@joanneum.at
<b>F. Fan</b>	Univ. Dalian (China)	Fanfengxiang@yahoo.com
<b>P. Grzybek</b>	Univ. Graz (Austria)	peter.grzybek@uni-graz.at
<b>L. Hřebíček</b>	Akad. d. W. Prag (Czech Republik)	ludek.hrebicek@seznam.cz
<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>G. Wimmer</b>	Univ. Bratislava (Slovakia)	wimmer@mat.savba.sk

**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. 25 (2013), Lüdenscheid: RAM-Verlag, 2013. 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 25 (2013)

**ISSN 2625-8226**

## Contents Glottometrics 25

### **Ramon Ferrer-i-Cancho**

Hubiness, length and crossings in syntactic dependencies 1-21

**Abstract.** Here tree dependency structures are studied from three different perspectives: their degree variance (hubiness), the mean dependency length and the number of dependency crossings. Bounds that reveal pairwise dependencies among these three metrics are derived. Hubiness (the variance of degrees) plays a central role: the mean dependency length is bounded below by hubiness while the number of crossings is bounded above by hubiness. Our findings suggest that the online memory cost of a sentence might be determined not just by the ordering of words but also by the hubiness of the underlying structure. The 2<sup>nd</sup> moment of degree plays a crucial role that is reminiscent of its role in large complex networks.

### **Karl-Heinz Best**

Zur Verslänge im Altisländischen 22-29

**Abstract.** In this contribution the distribution of word numbers in poetic texts in the Old Icelandic *Edda* is tested. The displaced binomial distribution seems to be the best model. But there are five cases in which the empirical findings deviate from this model. Four times other models could be fitted successfully.

### **Róisín Knight**

Laws governing rank frequency and stratification in English texts 30-42

**Abstract.** There are several laws that attempt to capture the regularities that seem to exist in the frequency structure of texts, by expressing the relationship between frequency and rank of words in a text. Within this field of research it has been found that stratification exists on many different levels, and the hypothesis proposed by Popescu, Altmann and Köhler (2010) allows for this to be explored further. This paper will use the method suggested by Popescu, Čech and Altmann (2011), to consider the presence of stratification in a new data set of English texts. Due to the fact that the study of this topic is a relatively new pursuit within linguistics, there is much confusion surrounding the question of what specific linguistic factors cause stratification. This paper attempts to answer this question, and tests Popescu, Mačutek and Altmann's (2009) theory that the number of strata in a text relates to the number of actors. However the results show that none of the texts studied, either containing a single actor or multiple actors, were found to be monostratal. Therefore the cause of stratification in texts is currently unknown, and until the mathematical representations of strata are able to shed light on this their application is limited.

### **Ioan-Iovitz Popescu, Peter Zörnig, Gabriel Altmann**

Arc length, vocabulary richness and text size 43-53

**Abstract.** The article describes the behaviour of arc length computed from the ranked frequencies of some text units and strives for constructing an indicator which is

independent of text size. Such an indicator may be used for text characterization, text comparison and classification, even for language comparisons.

### **Peter Zörnig**

A continuous model for the distances  
between coextensive words in a text 54-68

### **Olha Pavlyshenko**

The clustering of texts of English fiction  
in the vector space of semantic fields 69-84

**Abstract.** In this paper the lexical-semantic groups of verbs in texts of English fiction have been considered. It is shown that the frequency distribution of lexical-semantic groups of verbs in texts of English fiction makes it possible to characterize the lexical-semantic structure of author's idiolect. The strongest characterization potential is concealed in the frequency distribution of lexical-semantic fields that are formed by the verbs. The area of high-frequency words contains the words of nominative, stylistically neutral type, and the area of author's idiolect is located on the periphery of the lexical-semantic field. The constants of semantic distances that characterize the area of author's idiolect in the structure of lexical-semantic fields do not depend on the quantity and quality of authors' texts and represent the fundamental lexical-semantic regularities of author's style.

### **Gabriel Altmann, Ioan-Iovitz Popescu, Dan Zotta**

Stratification in texts 85-93

**Abstract.** Stratification in texts is a process analogous to those in nature and culture. Though one cannot identify the individual strata in every case, it is possible to show the rise of this phenomenon in mathematical terms and apply the resulting formulas to examples from textology and music. It allows also to study the evolution of a writer, text sort, language or music.

## **Report**

### **Martina Benešová, Dan Faltýnek**

Research activities at the Department of General Linguistics  
of the Philosophical Faculty of Palacký University in Olomouc,  
Czech Republic 94-96

## **Review**

**James W. Pennebaker** (2011): *The Secret Life of Pronouns -  
What our Words Say about Us*. New York: Bloomsbury Press  
Reviewed by **Jingqi Yan**, Zhejiang University. 97-101