

# **Glottometrics 24 2012**

**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. 24 (2012), Lüdenscheid: RAM-Verlag, 2012. 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 24 (2012)

**ISSN 2625-8226**

# Glottometrics 24, 2012

**Karl-Heinz Best**

*Diversifikation der starken Verben im Deutschen*

1-4

**Abstract.** In this paper the 1-displaced geometric distribution has been fitted to the ranked distribution of classes of the strong verbs in German. The classes are defined by the different vocalic alternations of the verbs. This way the paper brings a further corroboration of the hypothesis that diversification processes abide by laws.

**Vadim S. Baevskij**

*Non-traditional approach to the study of the rhythmic of Russian verse*

5-11

**Abstract.** The paper suggests a more precise method of analysing the iambic and trochaic rhythms of Russian verse, as compared with the traditional one – from Andrey Bely to A.N. Kolmogorov and M.L. Gasparov inclusive. The method described is based on the linguistic ideas proposed by A.A. Potebnya.

**Zhao Xiaodong**

*Vocabulary growth of content words of ESP and General English.*

*A contrastive study based on CMTE and BNC*

12-24

**Abstract.** This paper, based on Corpus of Maritime Transportation English (CMTE) and sampled British National Corpus (SBNC), employs FoxPro programs and SPSS analysis to study the dynamic growth patterns of words and content words of English for Specific Purposes (ESP) and general English at 4000-word intervals. Then it is tested in the paper whether Brunet's model can provide a good fit for the overall vocabulary growth of CMTE and SBNC, and whether this model is fit for describing the relationship between the vocabulary of content words and text length with the increase of tokens at 4000-word intervals. Lastly, the 95% confidence interval for content words in CMTE and general English is calculated.

Results of the study show that with the increase of cumulative tokens CMTE and SBNC exhibit a similar pattern of overall vocabulary increase, and the vocabulary increase curves of content words in the two corpora are also quite similar, with nouns increasing more rapidly than other content words. The difference is in SBNC overall number of words and content words increase more and more rapidly than those of CMTE, which means general English has greater vocabulary sizes of nouns, verbs, adjectives and adverbs. In addition, the vocabulary increase rate of SBNC tends to level with that of CMTE when the cumulative number of tokens reaches about 680000; the net increase of verbs in SBNC tends to slow down after the number of tokens reaches 350000. And in both general English and ESP, there is more inter-textual verb repetition, but less inter-textual adjective repetition. SPSS regression analyses show that Brunet's model can capture the vocabulary growth patterns of CMTE and the growth patterns of content words in CMTE and SBNC as well, with the determination coefficients ( $R^2$ ) all close to 1.

**Ioan-Iovitz Popescu, Zuzana Martináková-Rendeková, Gabriel Altmann**

*Stratification in musical texts based on rank-frequency distribution*

*of tone pitches*

25-40

**Abstract.** In this study we investigate the stratification in musical texts based on rank-frequency distribution of tone pitches. Preliminary investigations show that there are some similarities

between music and language and that pitches play the same semiotic role as phonemes or graphemes, occupying approximately the same “corridor” in the ternary plot. In this case we try to apply the stratificational approach and to investigate possible strata composed of pitches having different functions.

**Arjuna Tuzzi, Ioan-Iovitz Popescu, Peter Zörnig, Gabriel Altmann**

*Aspects of the behaviour of parts-of-speech in Italian texts*

41-69

**Abstract.** The present article is a continuation of the analysis performed in Tuzzi, Popescu, Altmann (2010). Here the parts-of-speech have been scrutinized. Their rank-frequency distributions have been characterized using the Repeat rate, the Entropy and Ord’s criterion. The ranking of POS with individual Italian presidents has been used to characterize the homogeneity individually and as a whole of 63 texts using Kendall’s concordance coefficient. There is high concordance. The last aspect is the computation of distances between identical parts-of-speech which yield a very unique picture represented by the Zipf-Alekseev function.

**Emmerich Kelih, Peter Zörnig**

*Models of morph lengths: Discrete and continuous approaches*

70-78

**Abstract.** We discuss a discrete and a continuous approach for modelling the distribution of morph length. The proposed models (1-displaced extended binomial distribution, beta function) have been successfully fitted to Spanish, Russian and Slovenian data records. It can be shown that the models are suitable for modelling word form types as well as word form tokens.

**Wang Hua**

*Length and complexity of NPs in Written English*

79-87

**Abstract.** This paper investigates the length distribution and the complexity of NPs of written English using the written section of the ICE-GB corpus as the data source. The results show that NPs have very complex patterns. The distribution of NPs and their patterns is affected by NP length. Such relationships can be exactly described with mathematic models.

## **History of Quantitative Linguistics**

**Peter Grzybek**

*Harry Dexter Kitson (1886-1959)*

88-94

## **Book Reviews**

**Gordana Đuraš:** *Generalized Poisson models for word length frequencies in texts of Slavic languages.* Diss. University of Graz.

Reviewed by **E. Nemcová**

95-96

**Emmerich Kelih:** *Die Silbe in slawischen Sprachen.* München - Berlin - Washington D.C.: Sagner. 188 pp.

Reviewed by **E. Nemcová**

97-99