

# **Glottometrics 11**

## **2005**

**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>P. Grzybek</b>	Univ. Graz (Austria)	peter.grzybek@uni-graz.at
<b>A. Hardie</b>	Univ. Lancaster (England)	a.hardie@lancaster.ac.uk
<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>V. Kromer</b>	Univ. Novosibirsk (Russia)	kromer@newmail.ru
<b>O. Rottmann</b>	Univ. Bochum (Germany)	otto.rottmann@t-online.de
<b>A. Schulz</b>	Univ. Bochum (Germany)	reuter.schulz@t-online.de
<b>G. Wimmer</b>	Univ. Bratislava (Slovakia)	wimmer@mat.savba.sk
<b>A. Ziegler</b>	Univ. Graz Austria)	Arne.ziegler@uni-graz.at

**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. 11 (2005), Lüdenscheid: RAM-Verlag, 2005. 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 11 (2005)

**ISSN 2625-8226**

# Contents of Glottometrics 11

## **Ferrer i Cancho, Ramon; Servedio, Vito**

Can simple models explain Zipf's law for all exponents?

1-8

**Abstract.** H. Simon proposed a simple stochastic process for explaining Zipf's law for word frequencies. Here we introduce two similar generalizations of Simon's model that cover the same range of exponents as the standard Simon model. The mathematical approach followed minimizes the amount of mathematical background needed for deriving the exponent, compared to previous approaches to the standard Simon's model. Reviewing what is known from other simple explanations of Zipf's law, we conclude there is no single radically simple explanation covering the whole range of variation of the exponent of Zipf's law in humans. The meaningfulness of Zipf's law for word frequencies remains an open question.

## **Best, Karl-Heinz**

Zur Häufigkeit von Buchstaben, Leerzeichen  
und anderen Schriftzeichen in deutschen Texten

9-31

**Abstract.** This paper deals with frequency distributions of letters in German. As in other natural languages, letters of the alphabet do not appear equally often. On the contrary, each letter has its own characteristic frequency. Moreover, the frequencies of the letters differ from text to text. The purpose of this paper is to present some further frequency distributions of letters in German texts and text corpora and to show that they all follow the negative hypergeometric distribution. In some well-chosen cases, spaces and punctuation marks are considered, too.

## **Andersen, Simone**

Word length balance in texts:

Proportion constancy and word-chain-lengths in Proust's longest sentence

32-50

**Abstract.** Constancy phenomena in word length distributions of texts are demonstrated. The regularity of proportions is shown by intercorrelation of parts under differing kinds of partitioning. Length homogeneity  $r_A$  as a measure for the stability of the values of the distribution is developed. Balance number  $B$  refers to word-chains in line: Every  $B$  words the total number of syllables tends to be equal, indicated by decreased variance.

## **Mačutek, Ján**

Discrete distributions connected by partial summations

51-55

**Abstract.** In this article it will be shown that (1) usual discrete probability (mass) distributions can be transformed in (almost) any partial sums distribution, (2) discrete probability distributions can be presented as partial sums distributions if an adequate transformations can be found.

## **Best, Karl-Heinz**

Turzismen im Deutschen

56-63

**Abstract.** Many authors examined the influence of loanwords in German (cf. Best 2001; Körner 2004). This above all applies to borrowings from French, English, and Latin, and for some other languages as well (for ex. Greek, Italian, Spanish). But there are many languages like Turkish the influence of which on the German lexicon is nearly unknown. The present paper presents the development of Turkish borrowings in German and demonstrates that this process abides by the logistic law which in linguistics is known as Piotrowski Law.

## **Kantemir, Sergej; Levickij, Viktor**

Die statistische Analyse des semantischen Feldes der Farbbezeichnungen im Deutschen 64-97

**Abstract.** The syntagmatic and paradigmatic characteristics of colour terms in German were investigated. The use of statistical methods enabled by means of formal criteria to objectively reveal similarities and dissimilarities of syntagmatic and paradigmatic peculiarities of colour terms. The findings of the research are:

1. In modern German the colour characteristics of different things and phenomena are defined via diverse system of colour terms that is comprised of a number of frequent groups of words, and where the most frequently used colour terms such as *weiß*, *schwarz*, *rot*, *blau*, *grün*, *grau*, *braun* and *gelb* (67,77%) make up the nucleus of the semantic field, the rest of the words (32,23%) belong to its periphery.
2. By means of the basic colour terms modern German can structure colour perception of an individual and “verbalise” the semantic structure of colour phenomenon in national *linguistic image of the world*.

## **Ferrer i Cancho, Ramon**

Hidden communication aspects in the exponent of Zipf’s law

98-119

**Abstract.** Here we focus on communication systems following Zipf’s law. We study the relationship between the properties of those communication systems and the exponent of the law. We describe the properties of communication systems using quantitative measures of the semantic vagueness and the cost of word use. We try to reduce the precision and the economy of a communication system to a function of the exponent of Zipf’s law and the size of the communication system. Taking the exponent of the frequency spectrum, we show that semantic precision grows with the exponent whereas the cost of word use reaches a global minimum between 1.5 and 2 if the size of the communication system remains constant. We show that the exponent of Zipf’s law is a key aspect for knowing about the number of stimuli handled by a communication system and determining which of two systems is less vague or less expensive. We argue that the ideal exponent of Zipf’s law should be very slightly above 2.