

Glottometrics 4 2002

To Honor G. K. Zipf

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

G. Altmann	Univ. Bochum (Germany)	02351973070-0001@t-online.de
K.-H. Best	Univ. Göttingen (Germany)	kbest@gwdg.de
L. Hřebíček	Akad. d. W. Prag (Czech Republik)	ludek.hrebicek@seznam.cz
R. Köhler	Univ. Trier (Germany)	koehler@uni-trier.de
V. Kromer	Univ. Novosibirsk (Russia)	kromer@newmail.ru
O. Rottmann	Univ. Bochum (Germany)	otto.rottmann@t-online.de
A. Schulz	Univ. Bochum (Germany)	reuter.schulz@t-online.de
G. Wimmer	Univ. Bratislava (Slovakia)	wimmer@mat.savba.sk
A. Ziegler	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

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

Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Glottometrics. 4 (2002), Lüdenscheid: RAM-Verlag, 2002. 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 2 (2002)

ISSN 2625-8226

Contents of Glottometrics 4, 2000 (including abstracts)

To Honor G.K. Zipf

Balasubrahmanyam, V.K., Naranan, S.

Algorithmic information, complexity and Zipf's law

1

Abstract. Zipf's law of word frequencies for language discourses is established with statistical rigor. Data show a departure from Zipf's power law term at low frequencies. This is accounted by a modi-fying exponential term. Both arise naturally in a model for word frequencies based on Information Theory, algorithmic coding of a text preserving the symbol sequence, concepts from quantum statistical physics and computer science and extremum principles. The Optimum Meaning Preserving Code (OMPC) of the discourse is realized when word frequencies follow the Modified Power Law (MPL). The model predicts a variant of the MPL for the relative frequencies of a small fixed set of symbols such as letters, phonemes and grammatical words. The OMPC can be viewed as containing orderly and random parts. This leads us to a quantitative definition of complexity of a string (C) that tends to 0 for the extremes of 'all order' and 'all random' but is a maximum ($C = 1$) for a mixture of both (Gell-Mann). It is found that natural languages have maximum complexity. The uniqueness of Zipf's power law index ($\gamma = 2$) is shown to arise in four different ways, one of which depends on scale invariance characteristic of fractal structures. It is argued that random text models are unsuitable for natural languages. It is speculated that a drastic change in symbol frequency distribution starting from phrases is related to emergence of meaning and coherence of a discourse.

Roelcke, Thorsten

Efficiency of communication. A new concept of language economy

27

Abstract. George Kingsley Zipf is known not only as the "father" of language statistics or quantitative linguistics in general, but also as one of the first who discussed the phenomenon of linguistic economy in detail. The following discussion in linguistics and communication sciences shows a wide spread of more or less scientific grounded concepts. This great conceptual diversity however disturbs the scientific discussion further on. Hence in the following contribution a new concept of language economy that fulfils holistic (and atomistic) requirements will be shown.

Schroeder, Manfred

Power laws: from *Alvarez* to *Zipf*

39

Abstract. This article is a reprint of a chapter from M. Schroeder, *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* (33-38), New York: Freeman 1991, published here with the kind permission of the author and the Freeman Publishing House. The book, very popular among linguists, makes us realize that we are not alone in the universe of sciences but have joined with all other disciplines at least due to the omnipotent power law which, almost everywhere, bears the name of the linguist G.K. Zipf. At the same time it shows us how physicists look at language.

Wheeler, Eric, S.

Zipf's law and why it works everywhere

45

Abstract. Zipf's law is a consequence of independently categorizing items, and rank ordering the categories. Therefore, it can be applied to almost anything. Testing methods on random input helps us see what is the artifact of the method rather than the property of the subject matter.

Debowski, Lukasz

Zipf's law against the text size: a half-rational model

49

Abstract. In this article, we consider Zipf-Mandelbrot law as applied to texts in natural languages. We present a simple model of dependence of the law on the text size, which is featured by variable power-law tail and constant ratio of the most frequent words. As a result we derive several closed formulas, which accord with empirical data qualitatively and partially quantitatively. For example, there appears to be a minimal length of literary texts equal to ≈ 159 word tokens for English.

Kornai, András

How many words are there?

61

Abstract. The commonsensical assumption that any language has only finitely many words is shown to be false by a combination of formal and empirical arguments. Zipf's Law and related formulas are investigated and a more complex model is offered.

Montemurro, Marcelo A., Zanette, D.Frequency-rank distribution of words in large text samples:
phenomenology and models

87

Abstract. In this paper we revisit Zipf's law in the context of linguistics. The deviations from the original simple power law are analysed and a dynamic model for text generation is proposed whose parameters can be associated with some structural features of languages. Furthermore, for the case of large corpora a novel phenomenology is disclosed. In this case a quantitative description of all the scaling regimes is possible by considering the family of solutions of a single first order differential equation.