Glottometrics 5
2002

To Honor G. K. Zipf

RAM-Verlag

ISSN 2625-8226

Glottometrics

Glottometrics ist eine unregelmäßig erscheinende Zeitdchrift (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

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. 5 (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 5 (2002) **ISSN 2625-8226**

Contents of

Glottometrics 5, 2002

(To honor G. K. Zipf)

Abstract. In this paper we consider the possibilities of known Zipf-Mandelbrot canonical law modifications. The proposed modifications explain the behavior of the right tail of the distribution and the presence of a deflection in the central part of the distribution (a crater). It is shown that the average word information load is invariant to the sample heterogeneity and that the proposed usage measure "places" the words more correctly with regard to their "importance".

Abstract. At the 100th anniversary of the birth of George Kingsley Zipf, one striking fact about the statistical regularity that bears his name, Zipf's law, is that it seems to appear everywhere. We may ask these questions related to the ubiquity of Zipf's law: Is there a rigorous test in fitting real data to Zipf's law? In how many forms does Zipf's law appear? In which fields are the data sets claiming to exhibit Zipf's law?

Abstract. This article starts with Zipf's (1949) "Tool Analogy", where the artisan arranges and re-designs his tools in a way minimizing his total work; as a result, more frequently used tools tend to be nearer to him (better accessible), smaller and multifunctional. We then argue that short distance, small size and multifunctionality reflect not only a high overall relative frequency of usage, but in particular a high frequency of usage in the first steps of a variety of complex working procedures. Tool order – word order? This extended Tool Analogy fits to the tendency of more frequent words to obtain initial positions in frozen binomials (Fenk-Oczlon 1989) and the new finding (Fenk & Fenk-Oczlon 2002a,b) that the short, frequent and multifunctional function words tend to concentrate in the first part of sentences.

Abstract. The functional structure of human language networks in the brain could be revealed in an indirect way by measurements in the abstraction level of words. The result is a very large deterministic graph or network, respectively, which was unknown in mathematics up to now. The whole network can only be represented in a matrix. Following Shannon's theory, it displays optimum properties for information processing (maximum entropy). The structure of the network can be subdivided by introducing word classes with increasing magnitudes which could contribute to an understanding of the biological generation of networks. The hypothesis is that such facts are valid for all natural languages. Differences will exist only in the individual distribution of matrix dots. That means, speaking precisely, that every language has a distinct individual network structure of its own. Surprisingly it can be shown that the superior general type of the universal network structure can be generated by statistical experiments. The properties of these networks were compared with those of natural language networks which are definitely not statistical. Finally the enigma of the famous Zipf-diagram can be disclosed by observing networks and text paths which run inside of them along existing connections from node to node. A staircase curve emerges, which is a better description of reality than a smoothed power law. All this can be repeated by experiments, which means that eventually we found a transition from descriptive to constructive science. Therefore the new ideas could be applied immediately also in technology.
Certainly the basic biological language structure arose a long time ago. Later on the typical patterns of the network connections for different language families should have evolved separately and were almost certainly accompanied by optimization processes for maximum entropy. Nowadays the details of the connection patterns for any language have to be learned anew by every child, and in this process, unusual alterations are not allowed by its language community.

Abstract. First, the status of Zipf(-Menzerath)'s Law and its criticisms are discussed, and the application of power law models, particularly in linguistics, is supported from a general point of view. The following sections, empirical studies on dependencies are conducted which test the Zipf-Mandelbrot Law, other power law models (Menzerath-Altmann's Law, the length-frequency dependency), and the word length distribution on data from Hungarian (a text and a dictionary).

Abstract. According to a widespread conception, quantitative linguistics will eventually be able to explain empirical quantitative findings (such as Zipf's Law) by deriving them from highly general stochastic linguistic 'laws' that are assumed to be part of a general theory of human language (cf. Best (1999) for a summary of possible theoretical positions). Due to their formal proximity to methods used in the so-called exact sciences, theoretical explanations of this kind

are assumed to be superior to the supposedly descriptive-only approaches of linguistic structuralism and its successors. In this paper I shall try to argue that on close inspection such claims turn out to be highly problematic, both on linguistic and on science-theoretical grounds.

Abstract. Evidence for the truth in George Kingsley Zipf's Principle of Least Effort can be found in the deep attraction we have for anything promising us an easier route. The selling point of virtually all technology is the promise of new means to ease. But, beneath the vast glittering surface in the sea of hype runs a dissipative current. With so many things capitalizing on our aversion to effort, individually and collectively, as a species we're losing it because we're not using it. Seeding that awareness is the first step in reversing the flow.

Abstract. The analysis shows that there is no intrinsic oscillation in the relation between frequency and length of words. The rise of oscillation is caused by using moving averages for smoothing the extremely dispersed data.