Glottometrics 15
2007

RAM-Verlag

Glottometrics

Herausgeber – Editors

Contents
of
Glottometrics 15

Abstract. This paper investigates probability distributions of dependency distances in six texts extracted from a Chinese dependency treebank. The fitting results reveal that the investigated distribution can be well captured by the right truncated Zeta distribution. In order to restrict the model only to natural language, two samples with randomly generated governors are investigated. One of them can be described e.g. by the Hyperpoisson distribution, the other satisfies the Zeta distribution. The paper also presents a study on sequential plot and mean dependency distance of six texts with three analyses (syntactic, and two random). Of these three analyses, syntactic analysis has a minimum (mean) dependency distance.

Abstract. The history of the German language is a history depicting the influence of foreign languages on German, as has been portrayed in different publications on the influence of the English, French, and Italian languages. The influence of other modern languages, among them the Russian language, has not been analysed to a great extent. This paper, based on gained data, intends to determine whether the Piotrowski-Law applies to the process of word-borrowing from Russian into German.

Abstract. The purpose of this paper is to present some further evidence for the validity of the logistic law in the development of the dictionary. To this end we test some data on the increase of terms and signs in a technical language presented by Warner (2007).

Abstract. The lognormal distribution had long been thought to be the most appropriate probability distribution for Japanese sentence length distributions. Yet this view had been supported only by few researches with sparse sampling data and reasoning contradicting language reality. In order to show a more realistic approach, we analyzed a substantial number of samples. At first, 150 essays and short stories were drawn as a random sample, out of which any pieces of writing whose length was either less than 100 or more than 1000 sentences were excluded. As a result, 113 pieces remained as sample texts. We also paid attention to the kinds of sentences, separating those of dialogue from narrative ones. From each one of these 113 sample texts, three sentence length frequency distributions were acquired – the first one for a complete text, the second one for the collection of direct speech in the same text, and the third one for all the narrative parts excluding direct speech above. The results completely overturn the long-standing belief, proving that a lognormal distribution – which has been computed but will not be shown here – can never be well applied to Japanese sentence length

distributions. Our new findings indicate that in place of this lognormal distribution, the Hyperpascal distribution maintains an excellent goodness of fit.

Ján Mačutek, Ioan-Iovitz Popescu, Gabriel Altmann
Confidence intervals and tests for the h-point and related text characteristics   45-52

Abstract. Confidence intervals and tests for recently introduced text characteristics (the *h*-point and its relatives) are derived.

Reginald Smith
Investigation of the Zipf-plot of the extinct Meroitic language   53-61

Abstract: The ancient and extinct language Meroitic is investigated using Zipf's Law. In particular, since Meroitic is still undeciphered, the Zipf law analysis allows us to assess the quality of current texts and possible avenues for future investigation using statistical techniques.

Reinhard Köhler, Reinhard Rapp
A psycholinguistic application of synergetic linguistics   62-70

Abstract: The paper presents a new attempt to analyse the relationship between word familiarity and word frequency within the framework of synergetic linguistics. Whereas in psychology it is customary to apply correlational analyses to such questions the current paper sets up a functional model and tests it on empirical data from two large corpora and a psycholinguistic database.

Ioan-Iovitz Popescu, Gabriel Altmann
Writer´s view of text generation   71-81

Abstract: Generally, a "writer's view", defined by the angle between the ends of the word rank-frequency distribution, as seen from the *h*-point, should be limited in the interval [$\pi/2$, π]. However, as shown in the present paper with 176 texts from 20 languages, actually the lower limit appears to be the golden number $\varphi = 1.618...$, rather than $\pi/2 = 1.57...$

Peter Grzybek
On the systematic and system-based study of grapheme frequencies:
a re-analysis of German letter frequencies   82-91

Abstract: This study looks at the theoretical modeling of letter frequencies. Based on recent findings demonstrating the negative hypergeometric function to be an adequate model, a re-analysis of German data reported by Best (2005) is conducted, concentrating on a detailed examination of parameter behavior. It is shown that all parameters of this distribution behave regularly, if the analysis is based on the system's inventory size, rather than on the class of items occurring in the given sample. Directions for future research are pointed out, particularly involving factors influencing parameter values.

History of Quantitative Linguistics   92-100

Karl-Heinz Best, Gabriel Altmann
XXX. Gustav Herdan (1897-1968)   92-96

Emmerich Kelih
XXXI. B.I. Jarcho as a pioneer of the exact study of literature   96-100