Glottometrics 43
2018

In Remembrance of Fengxiang Fan, 1950–2018

A Pioneer of Quantitative Linguistics in China

RAM-Verlag

ISSN 1617-8351
e-ISSN 2625-8226

Glottometrics

Indexed in ESCI by Thomson Reuters and SCOPUS by Elsevier

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

External Academic Peers for Glottometrics

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. 43 (2018), Lüdenscheid: RAM-Verlag, 2018. 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 43 (2018)
online/ e-version ISSN 2625-8226 (print version ISSN 1617-8351)

Glottometrics 43, 2018

Abstracts

Alexei Vasilev, Ilona Vasileva

Text Length and Vocabulary Size: Case of the Ukrainian Writer Ivan Franko

Abstract. In the paper we study how the vocabulary size depends on the text length for the literary works of the famous Ukrainian writer Ivan Franko. We propose two models that explain growth of the vocabulary size due to increasing of the text length. In the core of the models are differential equations which allow us to obtain approximation functions. These functions contain phenomenological parameters whose values we estimate with the measured data.

Peter Zörnig, Michal Místecký

Quantifying the Importance of Stylometric Indicators: A Principal Component Approach to Czech Sonnets 1

Abstract. We apply multivariate statistical analysis to stylometric indicators, calculated for sonnets of a prominent Czech poet, J.S. Machar. The study makes it possible to quantify the importance of linguistic indicators and to partition the sonnets into stylistic groups. Moreover, it introduces the index of weighted occurrence, which is a way for a quantitative assessment of the importance of the studied properties. The procedure applied in the paper may be of universal use.

Tomas Melka

Stylistic study of *Omnilingual* by H. Beam Piper

Abstract: In order to capture stylistic features of H. Beam Piper's classical story "*Omnilingual*" (1957) a number of quantitative measures are drawn in. As the aid of such measures is enlisted, various possibilities present themselves. We shall restrict ourselves to a small choice of possible descriptions, focused on the vocabulary richness and the time-structured writing sequence.

This Piper-esque writing has entered the records of the sci-fi prose for the "Martian" periodic table of elements, being synonymous with a scientific "Rosetta-like stone" in the decipherment area. The work, while having a search potential in text analysis and stylistics, may incidentally add some luster to the validity of the science as a communicative channel in non-conventional circumstances.

Adam King

The Lexicon and the Noisy Channel: Words are shaped to avoid confusion

Abstract. Language exists in a noisy channel and so an optimized lexicon should be structured to avoid possible confusion among words. For a code to be optimal in a noisy channel, it should maximize the mutual information between what is sent and what is received through noise. Using

confusion matrices for English phonemes, this article investigates the extent to which the English lexicon is structured under pressures of a noisy channel. I find that the relative frequency values and phonological make up of English words cause the lexicon to be less likely to suffer confusion than a comparable baseline. I discuss the results with respect to the growing body of literature on the lexicon as an optimal code.

Michal Místecký, Jiang Yang, Gabriel Altmann

Belza-chain Analysis: Weighting Elements

Abstract. In the article, it will be shown that weighting of elements of Belza-chains is possible using a classification of relations of elements and the main words. Uninterrupted sequences of sentences/lines form a chain, which has a main word; the other elements are some lexical or grammatical references. By counting the number of different classes, one obtains a possible frequency ranking which can be modelled. The main model is the Zipf-Alekseev function, but one can also employ the Lorentzian function. The analysis can be of use in multifarious investigations of styles.

Sergey Andreev, Giuseppe G. A. Celano, Jiang Yang, Gabriel Altmann

Some Properties of Polysemy

Abstract. In the article, the polysemy of individual words in some poems (English, Chinese, Russian, German, Italian and Slovak) is examined. Polysemy has its distribution, one can compute quantitative motifs omitting the line boundaries, positional representation, and a number of other problems mentioned below. The results can be modeled. We propose three models but the examination is not finished. A number of texts in every language are necessary, in order to show the differences between text types, authors, languages. It is rather a program for further research.

History

Yaqin Wang, Haitao Liu

In Remembrance of Fengxiang Fan, 1950–2018
A Pioneer of Quantitative Linguistics in China