

# **Glottometrics 44 2019**

**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 **Druck-version** 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>S. Andreev</b>	Univ. Smolensk (Russia)	smol.an@mail.ru
<b>K.-H. Best</b>	Univ. Göttingen (Germany)	kbest@gwdg.de
<b>R. Čech</b>	Univ. Ostrava (Czech Republic)	cechradek@gmail.com
<b>E. Kelih</b>	Univ. Vienna (Austria)	emmerich.kelih@univie.ac.at
<b>R. Köhler</b>	Univ. Trier (Germany)	koehler@uni-trier.de
<b>H. Liu</b>	Univ. Zhejiang (China)	lhtzju@gmail.com
<b>J. Mačutek</b>	Univ. Bratislava (Slovakia)	jmacutek@yahoo.com
<b>A. Mehler</b>	Univ. Frankfurt (Germany)	amehler@em.uni-frankfurt.de
<b>M. Místecký</b>	Univ. Ostrava (Czech Republic)	MMistecky@seznam.cz
<b>G. Wimmer</b>	Univ. Bratislava (Slovakia)	wimmer@mat.savba.sk
<b>P. Zörnig</b>	Univ. Brasilia (Brasilia)	peter@unb.br

## External Academic Peers for Glottometrics

### Prof. Dr. Haruko Sanada

Rissho University, Tokyo, Japan (<http://www.ris.ac.jp/en/>);

Link to Prof. Dr. Sanada: <http://researchmap.jp/read0128740/?lang=english>; <mailto:hsanada@ris.ac.jp>

### Prof. Dr. Thorsten Roelcke

TU Berlin, Berlin, Germany ( <http://www.tu-berlin.de/> )

Link to Prof. Dr. Roelcke: [http://www.daf.tu-](http://www.daf.tu-berlin.de/menue/deutsch_als_fremd_und_fachsprache/mitarbeiter/professoren_und_pds/prof_dr_thorst)

[berlin.de/menue/deutsch\\_als\\_fremd\\_und\\_fachsprache/mitarbeiter/professoren\\_und\\_pds/prof\\_dr\\_thorst](http://www.daf.tu-berlin.de/menue/deutsch_als_fremd_und_fachsprache/mitarbeiter/professoren_und_pds/prof_dr_thorst)

[en\\_roelcke](mailto:Thosten.Roelcke@tu-berlin.de)  
[mailto:Thosten.Roelcke \(roelcke@tu-berlin.de\)](mailto:Thosten.Roelcke@tu-berlin.de)

**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. 44 (2019), Lüdenscheid: RAM-Verlag, 2019. 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 44 (2019)  
**online/ e-version ISSN 2625-8226 (print version ISSN 1617-8351)**

# Glottometrics 44, 201

## Abstracts

**Francesc Reina, Irene Castellón, Lluís Padró**

Towards the Prepositional Meaning via Machine Learning:  
A Case Study of Spanish Grammar

**Abstract.** Is it possible to identify or measure prepositional meaning? In our article we review a particular case of semantic universe, the verbs of movement in Spanish. In this context, we try to answer positively the initial question and validate a method. From the selection of a corpus of 71,206 prepositional phrases in Spanish, where three prepositions – *a*, *hacia* and *hasta* – are distributed, we proceed to verify the hypothesis about the semantic gradualness of the prepositions (HGSS). Applying tools of the field of machine learning, we establish a series of groupings that are compared with the hand annotation classification. The results are statistically relevant insofar as they confirm our initial hypothesis

**Guoqiang Zhang, Haitao Liu**

A Quantitative Analysis of English Variants Based on Dependency Treebanks

**Abstract.** Different English variants have been found to have different characteristics with respect to the aspects of pronunciation, vocabulary, spelling, and grammar, but there is a dearth of research from the perspective of sentence length and syntactic dependency. Therefore, the present research studies these two aspects based on 12 self-building academic dependency treebanks and finds that: (1) in terms of sentence length and syntactic dependency, different variants of English manifest great similarities; (2) regarding the distribution of dependency distance of English variants, the parameters of right-truncated modified Zipf-Alekseev model can provide evidence for the different stages of English globalization; (3) in addition, their syntactic dependencies are little affected by their respective native languages, suggesting the important roles of grammar and cognition in the syntactic constraint. The present research provides a new perspective to study English variants, and its conclusion is hopefully expected to enrich the current research theory.

**Xiaowen Zhang, Yunhua Qu, Zhiwei Feng**

A Corpus-Based Study on the Diachronic Relationship  
between the Contemporary American English Present Perfect  
and Simple Past Across Registers

**Abstract.** The relationship between the diachronic change of the present perfect (PP) and the past simple tense (SP) in English has always been an important, but still puzzling subject of studies on the English aspect-tense system, because the two constructions are both related to past-time reference, and the distinction between the two is not clear-cut in many English varieties. Even contexts labeled by temporal adverbials – like *yesterday*, which tends to be used with the SP, or *since*, which is usually combined with the PP – have become increasingly compatible with the other construction. Therefore, it is assumed that the diachronic change of the PP or the SP should not be studied individually, as is

done in many previous studies, but observed from a broader perspective – the competition between the two.

This study, using the largest and most balanced American English corpus, COCA (The Corpus of Contemporary American English), aims to investigate the diachronic change patterns of the PP and the SP, as well as to detect their relationship, in five different registers (Spoken, Fiction, Magazines, Newspapers, and the Academic one). Findings show that the diachronic change of the PP or the SP is closely related, either negatively (i.e. competitively) or positively in the entire corpus, except for the Academic register. To be more specific, the development of the PP and the SP are in competition in COCA and the Spoken register, while positively related in Fiction, Magazine, and Newspaper. No statistically significant correlation is found in the Academic one.

This research adds the dimension of register into the study of the diachronic change of the English tense-aspect system, adopting relatively convincing statistical methods to reveal a panorama about the relationship between the development of the PP and the SP in contemporary American English.

**Aiyun Wei, Haitao Liu**

### Typological Features of Zhuang from the Perspective of Word Frequency Distribution

**Abstract.** Investigating lexical features with statistical methods has always been a key object of quantitative linguistic research. However, though Zhuang is the mother tongue of the minority with the largest population in China, its lexical features have attracted little attention from the researchers employing quantitative means. Based on a corpus (CZL) of over 500,000 tokens of the Zhuang language, this study addresses the features of word frequency distribution of Zhuang. The results show that Zhuang shares the universal feature of other tested languages in that its word frequency distribution abides by the Zipf's Law and the "Least Effort Principle". The study also tests the word frequency distribution of Zhuang texts of different genres, which shows that for different genres, the values of some parameters, such as  $b$ , are different. Moreover, in order to test whether Zhuang language has any distinctive or typological features in word frequency distribution, the values of the  $h$ -point and  $a$ -index of the texts in CZL are computed as well. It is found that the two indexes are effective in distinguishing Zhuang from other languages, and the position of Zhuang on the analytism-synthetism continuum proposed by Popescu is close to those of the Polynesian language family, which may be helpful for intersubjective placement of Zhuang into a language group. This study would open a new perspective in the statistical lexical research of Zhuang language and present a "new" corroborated language with respect to the laws in quantitative linguistics.

**Hong Ma, Haitao Liu**

### Probability Distribution of Causal Linguistic Features

**Abstract.** This study investigated the probability distribution of cause-effect language produced by four different groups of speakers, ESL and non-ESL high school and primary students. Results showed that, regardless of speakers' English proficiency, the probability distribution of causal linguistic features produced by native speakers fitted the right-truncated modified Zipf-Alekseev model, while this model did not capture the distribution of causal linguistic features identified in nonnative speakers' speech. The results of current research suggested that the right truncated modified Zipf-Alekseev model has the potential of differentiating between native and nonnative causal discourse. In addition, the fitting results shed light on relationship between parameters ( $a$  and  $b$ ) and students' proficiency levels as well as tentatively offer an alternative to systemic functional linguistic (SFL) method of evaluating students' causal language use.

## **Hanna Gnatchuk**

### Measuring Lexical Richness of the USA Presidents' Inauguration Speeches

**Abstract:** The present study deals with a quantitative and corpus-based study of inaugural speeches of the USA presidents in terms of lexical richness. The fifteen speeches have been analysed and the lexical richness of presidents has been compared by means of the one-way ANOVA. The results of the findings have shown us that the indices of the speeches by the USA presidents differ significantly. All the tests have been computed in statistical program R-Studio and in Python 3.

## **Panchanan Mohanty, Ioan-Iovitz Popescu, Gabriel Altmann**

### Script Complexity in Indian Languages

**Abstract.** The present article shows a method for computing script complexity known from other articles and presents the computing in seven Indian languages. At the same time it is a hint: if the scripts should be unified, one should choose the simplest script. The simplest script means the choice of the simplest letters, not the whole script. In print, letters are chosen because of the need for ornamentality; but for schools the simplest forms should be used.