

## History of Quantitative Linguistics

Since a historiography of quantitative linguistics does not exist as yet, we shall present in this column short statements on researchers, ideas and findings of the past – usually forgotten – in order to establish a tradition and to complete our knowledge of history. All contributions are to be sent to Peter Grzybek, [grzybek@uni-graz.at](mailto:grzybek@uni-graz.at).

### I. Viktor Jakovlevič Bunjakovskij

**Viktor Jakovlevič Bunjakovskij** was an important Russian mathematician (3.12.1804 - 30.11.1889), who played an important role in improving the level of mathematical education, as did other mathematicians of his time, such as M.S. Ostrogradskij or P.L. Čebyšev.- Bunjakovskij received his primary education in Moscow in the home of Count A.P. Tomasov, a friend of his father's who had died in 1809. From 1820, Bunjakovskij stayed abroad together with the latter's son; first he was in Coburg, then in Lausanne, finally in Paris, where he studied at the Sorbonne und the Collège de France, and heard lectures of Laplace, Fourier, and Poisson, among others. In 1824 Bunjakovskij received the bachelor's degree, in 1825, he was appointed doctor of mathematics by the Paris Faculté des sciences. Immediately after his return to Russia (1826), Bunjakovskij began to teach mathematics in Petersburg. From 1846-1859 he held lectures on analytical mechanics, differential and integral calculus, as well as on probability theory at Petersburg University. In 1828, Bunjakovskij was elected to an adjunct for pure mathematics of the Academy of Sciences. In 1830, he was elect to an extraordinary, and in 1836 to an ordinary member of the Academy. From 1864 until shortly before his death Bunjakovskij was Vice President of the Academy of Sciences. In 1883, Bunjakovskij, who mostly worked in the realm of number theory and probability theory, published his catalogue of scientific works (*Liste des travaux mathématiques de Victor Bouniakowsky*; SPb 1883), which points out 108 scholarly titles.

In 1847, Bunjakovskij published an article in the third volume (part II) of the journal *Sovremennik*, which was entitled: „On the possibility to apply determining measures of confidence to the results of some observing sciences, particularly statistics“. Irrespective of the importance this article has, in a historical perspective, with regard to establishing statistics as a methodological discipline in its own right, Bunjakovskij's article represents – not only for Russia – one of the earliest quotations where the possibility and reasonability of applying statistics to, among others, philological questions is being discussed.

### О ВОЗМОЖНОСТИ

ВВЕДЕНИЯ ОПРЕДЕЛИТЕЛЬНЫХЪ МѢРЪ ДОВѢРІЯ КЪ  
РЕЗУЛЬТАТАМЪ НѢКОТОРЫХЪ НАУКЪ НАБЛЮДАТЕЛЬНЫХЪ,  
И ПРЕИМУЩЕСТВЕННО СТАТИСТИКИ.

Thus, at the very ending of this article, Bunjakovskij writes:

“[...] It would be time now, to finish the present article; it may be allowed, however, in accordance with the analogy of the matter at stake, to add a couple of words with regard to another application of probability analysis, to which obviously no-one has ever before drawn the attention. The new application includes grammatical and etymological studies of a language, as well as comparative philology. However strange such studies may, at first sight, seem for a mathematical analysis, yet I am fully convinced that a broad field for strictly mathematical ruminations opens up before us. My claim is not so much based on more or less uncertain conjectures and suppositions but on a critical evaluation of the object, on a number of attempts I have already made, and some analytical formulae, which I have introduced in order to define the *numerical* probability of particular word formations. In this way one can, for example, approximately determine the measure of confidence of a given etymology, and, depending on the proximity of this number to 1 and its confidence, one has to judge about the proposed authenticity. This is not the place to discuss details of this matter to which I merely wanted to draw the attention. But in order to directly explain in which way corresponding studies may find their way into the realm of applied mathematics, I regard it to reasonable to briefly point out some of the numerical assertions and materials which need to be elaborated. If we talk about a language, we predominantly assume that we have its detailed *arithmetical description* or, if we will, its statistics, i.e. numerical assertions about the complete inventory of the words of that language, about the distribution of these words according to the parts of speech, about the number of letters, about the initial letters, the endings, etc. etc. Also general rules have to be named here, exceptions of various kinds, words undoubtedly stemming from different languages, and the like. This is the numerical material the strict analysis of which quite naturally demands mathematical considerations. If one has the relevant statistical data as to two or more languages, one can compare them in various aspects, and the results obtained take the status of an authority which philologists, at the present state of science, do not use to have available.

Of course the elaboration of what I have called the *statistics of language* is a very painstaking endeavor, and most probably philologists will tend to call this kind of effort almost wasted, because the assumed gain of the exactness of the conclusions about language does not pay for the loss of time. We do not take on the responsibility of solving the question to which degree such an assertion may be justified.

Maybe I will publish my theoretical studies on the matter which I have only mentioned here, at some other opportunity. However, as far as the practical applications of the general formulae are concerned, it will be necessary, given the general lack of detailed arithmetical data about languages, to confine oneself to a few examples. By the way, in order to devote the proper degree of completeness and soundness to such a work with regard to the philological conditions, it goes without saying that a mathematician has to establish direct contact with specialists in this field, more or less foreign to himself.”

Peter Grzybek

# Glottometrics

**Glottometrics** ist eine unregelmäßig erscheinende Zeitschrift 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, auf **CD-ROM** (in PDF Format) oder in **Buchform** bestellt werden

**Glottometrics** is a scientific journal for the quantitative research on language and text published at irregular intervals

**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**, obtained on **CD-ROM** (in PDF) or in form of **printed copies**

## Herausgeber – Editors

G. Altmann	<a href="mailto:02351973070-0001@t-online.de">02351973070-0001@t-online.de</a>
K.-H. Best	<a href="mailto:kbest@gwdg.de">kbest@gwdg.de</a>
P. Grzybek	<a href="mailto:grzybek@uni-graz.at">grzybek@uni-graz.at</a>
L. Hřebíček	<a href="mailto:hrebicek@orient.cas.cz">hrebicek@orient.cas.cz</a>
R. Köhler	<a href="mailto:koehler@uni-trier.de">koehler@uni-trier.de</a>
V. Kromer	<a href="mailto:applied@newmail.ru">applied@newmail.ru</a>
O. Rottmann	<a href="mailto:otto.rottmann@t-online.de">otto.rottmann@t-online.de</a>
A. Schulz	<a href="mailto:reuter.schulz@t-online.de">reuter.schulz@t-online.de</a>
G. Wimmer	<a href="mailto:wimmer@mat.savba.sk">wimmer@mat.savba.sk</a>
A. Ziegler	<a href="mailto:arneziegler@compuserve.de">arneziegler@compuserve.de</a>

**Bestellungen** der CD-ROM oder der gedruckten Form sind zu richten an  
**Orders** for CD-ROM's or printed copies to

RAM-Verlag [RAM-Verlag@t-online.de](mailto:RAM-Verlag@t-online.de)

**Herunterladen / Downloading:** <http://www.ram-verlag.de>

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Glottometrics. –6 (2003) –. – Lüdenscheid: RAM-Verl., 2003

Erscheint unregelmäßig. – Auch im Internet als elektronische Ressource unter der Adresse <http://www.ram-verlag.de> verfügbar.-

Bibliographische Deskription nach 6 (2003)

**ISSN 1617-8351**

# Contents

<b>Hřebíček, L.</b> Some aspects of the power law	1-8
<b>Best, K.-H.</b> Spracherwerb, Sprachwandel und Wortschatzwachstum in Texten. Zur Reichweite des Piotrowski-Gesetzes	9-34
<b>Wilson, A.</b> Word-length distribution in modern Welsh prose texts	35-39
<b>Dshurjuk, T.V., Levickij, V.W.</b> Satztypen und Satzlängen im Funktional- und Autorenstil	40-51
<b>Rottmann, O.</b> Word length in the Baltic languages – are they of the same type as the word lengths in the Slavic languages?	52-60
<b>Strauss, U., Altmann, G.</b> Age and polysemy of words	61-64
<b>Wheeler, E.S.</b> Multidimensional scaling to visualize text separation	65-69
<b>Jüngling, R., Altmann, G.</b> Python for linguistics?	70-82
<b>Popescu, Ioan-Iovitz</b> On a Zipf's Law extension to impact factors	83-93
<b>Project report</b> <b>Kelih, E., Grzybek, P., Stadlober, E.</b> Das Grazer Projekt zu Wortlängen(häufigkeiten)	94-102
<b>History of Quantitative Linguistics</b> I. V.J. Bunjakovskij (by P. Grzybek) II. B. Trnka – The first bibliography (by L. Uhlířová)	103-106
<b>Books received</b>	107