

- Resnik, P., & Yarowsky, D. (2000). Distinguishing systems and distinguishing senses: New evaluation methods for word sense disambiguation. *Natural Language Engineering*, 5(2), 113–133.
- Schütze, H. (1998). Automatic word sense discrimination. *Computational Linguistics*, 24(1), 97–123.
- Snyder, B., & Palmer, M. (2004). The English all words task. *Proceedings of Senseval-3: The 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text* (pp. 41–43). Barcelona, Spain.
- Véronis, J., & Ide, N. M. (1990). Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. *Proceedings of the International Conference on Computational Linguistics (COLING)* (pp. 588–594). Helsinki, Finland.
- Yarowsky, D. (1992). Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)* (pp. 454–460). Nantes, France.
- Yuret, D. (2004). Some experiments with a naïve Bayes WSD system. *Proceedings of Senseval-3: The 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text* (pp. 265–268). Barcelona, Spain.

BOOK REVIEW

Kelih, Emmerich. *Geschichte der Anwendung quantitativer Verfahren in der russischen Sprach- und Literaturwissenschaft*. Studien zur Slavistik, Bd. 19. Hamburg: Kovač, 2008. ISBN: 978-3-8300-3575-6, 346 pp.

Reviewed by Haitao Liu

School of International Studies, Zhejiang University, China

Quantitative linguistics (QL) is a branch of linguistics but it is difficult to find QL-related topics in comprehensive books or works on linguistics and its history. History, theory and methodology can be considered as the three foundation stones of a science. If students and researchers who are interested in QL cannot easily find the needed QL materials it is very harmful to the development of the science.

Compared with theory and methodology of QL, there is less material on QL history, but the situation is improving. For example, *Glottometrics* publishes some notes about QL history, and the first section of Köhler et al. (2005) also contains 10 articles on QL history. However, this is not enough to build a new science such as QL. We urgently need works that systematically treat the history of QL on an international and national scale. As Kelih's book is one such work, it is a pleasure to introduce it to the readers of this journal.

The goal of the work under review is the systematic research of the application of quantitative methods in Russian linguistics and literature science. It concerns a historical investigation of the past employment of quantitative means in these fields. The book consists of 10 chapters and a comprehensive list of references (pp. 283–338), which can be used as a guide for further research.

Considering the difficulty to make a historical study of QL which is not a popular and systemized science, the author sets up the objectives and

methods of his research in the first chapter (pp. 13–30). After investigating several important linguistic resources, the author found that quantitative methods are often ignored or marginalized in the history of linguistics and literature science. So, the author proposes to reconstruct the studied history rationally. The central starting point of the “rational reconstruction”, which came from the philosopher of science Imre Lakatos, is to represent the historical process of a science as the succession of rivaling research theories and research programs. Based on the philosophy of science, the author methodologically raises the following questions (p. 28):

- (1) Where do the determined “research programs” come from which rely on the application of statistical means?
- (2) Are the investigations based on deductive theories and empirically examinable (falsifiable) hypotheses?
- (3) Is there an emphasis relating to contents from the application of statistical means?
- (4) Is the use of statistical methods bound to linguistic or literature trends?
- (5) Which theoretical and methodological problems are given in the case of quantitative language and text analysis?
- (6) Which central knowledge or resource is produced from the application of statistical methods?

In order to discuss these questions with the reconstruction of the history of quantitative means in Russian linguistics and literature science, the author also considers two central factors, *viz.* “science-internal” (methods, hypotheses, results) and “science-external” (political influences, etc.) factors. This is understandable and necessary, because it is obvious that the political factor always plays an important role in all fields of any socialist country.

The second chapter (pp. 31–44) summarizes the works of V. A. Bunjakovskij, A. S. Budilovič, M. A. Dikarev, and the early quantitative verse analysis and frequency dictionaries. Strictly speaking, these works only can be regarded as a prehistory of the application of statistical means in linguistics and literature. However, this modest beginning also shows the following: philology (similar to medicine, sociology, economics etc.) was understood as a meaningful application field of statistics; the quantitative structure of language and text already covered a very broad spectrum, which contains frequency of letters and sounds and

comparisons among different languages, statistical verse analysis, frequency of metric forms, position of stresses and length of text units, etc.* Although this broad interest in quantitative linguistic and text characteristics were usually based on “simple” counting, people already partially defined the use of statistics in language and text analysis as an independent discipline with names such as “language statistics”, “linguistic statistics”, “dictionary statistics” and so on. So, this was a good beginning.

The third chapter (pp. 45–81) with the title “Statistics and linguistics” covers the time span from the first decade to the fourth decade of the 20th century. It seems to me that this chapter is one of the most important components in written history and it is worth listing its subtitles here: from stylometrics to probability; the Kazan linguistic school and statistical means; from sound to syntax; experimental phonetics and spreading of statistical methods; statistical analysis of diachronic processes; practical applications: construction of “economic” alphabets, language didactics and language training research; typography and stenography; statistics as bridge between linguistics and literature science; the first Russian monograph of language statistics; quantitative methods in the 1930s and 1940s.

Several interesting points can be summarized in this phase. At the beginning, two opposite directions of the application of statistical methods, which were represented by the works of N. A. Morozov and A. A. Markov, appeared: on the one hand professional statistics, which uses probability theory to approach useful empirical materials and test hypotheses on linguistic data, and on the other hand, the more or less dilettante style investigations, in which simple statistical methods are used, but with a lack of essential interpretations of the findings. In fact, the struggle between these two tendencies still exists today in all countries. Perhaps, it is an eternal task for researchers to decide how to locate themselves between these two extremes. The second substantial function of statistical analysis is seen at this time in the comparison of linguistic phenomena, whereby statistics is understood as adequate equipment, which is in the situation to compare several objects regarding their quantitative development. The author also shows that the sound level was located at the centre of the interest. This is a significant summary, because up to today, the sound level is still one of the most solid foundations for quantitative linguists. As remarked by the author, this can be interpreted in the sense of a system-stable language

characteristic, which gives a lasting idea of the quantitative structure of languages. The sound level is also the starting point for first investigations about language typology and text classification. The frequency of sounds goes thereby as a characteristic into action, which permits it to classify languages or texts, although similar text-typological investigations are also made on the basis of syntactic parameters (frequency and the length of syntactic constructions).

The fourth chapter (pp. 83–142) reconstructs the history of quantitative means in the Russian literature science from 1910 to the 1930. The contents can also be called a historical investigation of exact literature science in Russia. According to the author, the studies in this period can be arranged into three development stages. The first stage covers the establishment of statistical methods in Belyj's *Simvolizm* with which the studies are closely connected, both conceptually and methodologically. The second development stage represents a descriptively-oriented phase. The goal of these investigations ranges from the description of the verse and text to observing phenomena. The third stage of the history of quantitative means in verse and text analysis is the successful synthesis of statistics, formal text analysis and linguistics in the form of the exact literature science of B. I. Jarcho.

The fifth chapter (pp. 143–171) covers the period from the middle to the end of the 1950s. This is an important phase for establishing quantitative means in Russian linguistics and literature science. But, more interestingly, the author clearly presents how science-external (political and cultural, historical) and science-internal factors (contents, methods and objects of the statistical investigation) are closely linked with one another. The subtitles will reveal the secrets and complexity of the history: from dogmatism to progressive methodological consciousness; structuralism debates and the integration of quantitative means; conferences on statistical and mathematical linguistics; statistics, semiotics and cybernetics – reintegration of the exact literature science; quantitative methods in machine translation.

The sixth chapter (pp. 173–204) discusses and analyses some notable works on statistical linguistics, which were developed from the beginning to the end of the 1960s. An important characteristic of this phase is that methods of inference statistics were applied to linguistics. Another trait is the statistical means used for solving practical problems of the technical language analysis and synthesis. At the end of this chapter, the author also discusses science-political pressure on statistical methods.

The following chapter (pp. 205–227) summarizes the most important achievements of the research group “Statistika Reči”, which concerns a loose association of over 200 scientists from linguistics, computer science, statistics, mathematics and psychology, from the beginning of the 1960s into the 1990s, within the overall framework of language statistics. The members of the group were interested in the interdisciplinary study of linguistic structures with the help of statistical and quantitative methods. The significant contribution of the group to language statistics are:

- (1) “Mathematical linguistics” is defined as a discipline which covers both quantitative (statistics, probability calculation) and non-quantitative (algebra, differential calculus) methods of mathematics.
- (2) As a direct component of “Statistika Reči”, a new discipline, engineering linguistics, was created for the application of statistical methods in linguistics; it provides automated ways of gathering raw linguistic data for quantitative analysis.
- (3) The elaboration and extension of general objects for statistical analyses of the linguistic standard, which is introduced as statistically analyzable intermediate level between language system and speech.
- (4) The intensive and solid arguments with the methodical apparatus of statistics were meaningfully suitable for linguistic investigations.

Another original and innovative contribution of “Statistika Reči” is the measurement of the quantitative information structure of words and/or word forms. Based on these facts, the author considers that the work of “Statistika Reči” is a successful symbiosis of automated language analysis and the application of statistical methods in Russian linguistics.

According to the strict definition, the works mentioned above cannot be called quantitative linguistics in the modern sense. So, the eighth chapter (pp. 229–262) presents how language statistics has changed to quantitative linguistics in Russia. This evolution is closely connected with names such as J. A. Tuldava, M. V. Arapov and Ju. K. Orlov. The ninth chapter (pp. 263–270) gives the current tendencies. The last chapter (pp. 271–282) is the summary of the whole book.

In conclusion, the book focuses on the history of quantitative means in Russian linguistics and literature science. The central problem is the historical foundation and evolution of statistical methods in language and text analysis. Based on the amount of first-hand materials, the

author clearly reconstructs the history of quantitative means in Russian linguistics and literature science.

The Russians have made an important contribution to the development of QL and computational (or mathematical) linguistics, but it is not easy for researchers to access these results in Russian. The difficulty is not only in the language barrier, but also in the rarity of the material. For instance, Zhi (2009) is seen as one of the most comprehensive works on Russian linguistics, but it does not include any information about QL in Russia. Thus, Kelih's book is very welcome to all readers outside and inside Russia. It is not only useful to quantitative linguists, but is also an important reference for mathematical and computational linguists.

REFERENCES

- Köhler, R., Altmann, G., & Piotrowski, R. G. (Eds.) (2005). *Handbook of Quantitative Linguistics*. Berlin: de Gruyter.
- Zhi, Y. (2009). *General History of Russian Linguistics [Eluosi yuyanxue tongshi]*. Shanghai: Foreign Language Education Press.

Subscription Information

The *Journal of Quantitative Linguistics* (Print ISSN 0929-6174) is a peer-reviewed journal published four times a year (February, May, August and November) by Taylor & Francis, an Informa business, 4 Park Square, Milton Park, Abingdon, Oxfordshire OX14 4RN, UK.

Annual subscription, Volume 18, 2011

Institutional (print and online):	£374	US\$641	€511
Institutional (online only):	£337	US\$577	€460
Personal:	£141	US\$241	€191 (plus VAT where applicable)

Members of the International Quantitative Linguistics Association (IQLA) can subscribe at a reduced rate. Information about IQLA can be found at <http://www.iqla.org>. Members as well as individuals must declare that the subscription is for their own private use, that it will not replace any existing institutional subscription, and that it will not be put at the disposal of any library.

Dollar rates apply to subscribers in all countries except the UK and the Republic of Ireland where the pound sterling rate applies, and in the rest of Europe where the euro rate applies. Subscriptions purchased at the personal rate are strictly for personal, non-commercial use only. The reselling of personal subscriptions is prohibited. Personal subscriptions must be purchased with a personal cheque or credit card. Proof of personal status may be requested. For full information please visit the Journal's homepage at <http://www.tandf.co.uk/journals>

Back Issues

Taylor & Francis retains a three-year back issue stock of journals. Older volumes are held by our official stockists: Periodicals Service Company <http://www.periodicals.com/tandf.html>, 11 Main Street, Germantown, NY 12526, USA, to whom all orders and enquiries should be addressed. Telephone: +1 518 537 4700; Fax: +1 518 537 5899; Email: pse@periodicals.com

Offprints and Complimentary Issues

Corresponding authors will receive free online access to their article through our website (www.informaworld.com) and a complimentary copy of the issue containing their article. Reprints of articles published in *Journal of Quantitative Linguistics* can be purchased through Rightslink® when proofs are received or alternatively on our journals website. If you have any queries, please contact our reprints department at reprints@tandf.co.uk. Additional copies of the journal can be purchased at the author's preferential rate of £15.00 per copy.

Copyright and Permissions

It is a condition of publication that authors assign copyright or license the publication rights in their articles, including abstracts, to Taylor & Francis. This enables us to ensure full copyright protection and to disseminate the article, and of course the journal, to the widest possible readership in print and electronic formats as appropriate. Authors retain many rights under the Taylor & Francis rights policies, which can be found at www.informaworld.com/authors_journals_copyright_position. Authors are themselves responsible for obtaining permission to reproduce copyright material from other sources.

Taylor & Francis grants authorization for individuals to photocopy copyright material for private research use, on the sole basis that requests for such use are referred directly to the requestor's local Reproduction Rights Organization (RRO). The copyright fee is \$32 exclusive of any charge or fee levied. In order to contact your local RRO, please contact: International Federation of Reproduction Rights Organizations (IFRRO), rue du Prince Royal, 87, B-1050 Brussels, Belgium, e-mail: ifro@skynet.be; Copyright Clearance Center Inc., 222 Rosewood Drive, Danvers, MA 01923, USA, e-mail: info@copyright.com; Copyright Licensing Agency, 90 Tottenham Court Road, London, W1P 0LP, UK, e-mail: cla@cla.co.uk. This authorization does not extend to any other kind of copying by any means, in any form, for any purpose other than private research use.