

Iva Nebeská; Jarmila Panevová

Book review. Prague studies in mathematical linguistics 4

*Czechoslovak Mathematical Journal*, Vol. 24 (1974), No. 3, 507–509

Persistent URL: <http://dml.cz/dmlcz/101266>

## Terms of use:

© Institute of Mathematics AS CR, 1974

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://dml.cz>

BOOK REVIEWS

PRAGUE STUDIES IN MATHEMATICAL LINGUISTICS 4, Academia, Praha 1972, 254 pp., 65,— Kčs.

The series of PSML has been, since the appearance in 1966 of its first volume,<sup>1)</sup> the platform of Czechoslovak mathematical linguists; it includes contributions by linguists, mathematicians, and specialists in computational linguistics and information science. The fourth volume, the scientific editors of which are J. Horecký, P. Sgall and M. Těšitelová, is divided — similarly to the previous three — into two parts called Quantitative Linguistics, and Algebraic Linguistics, respectively.

In the first part, *M. Těšitelová* has concentrated the focus of her study on the important question of the choice of such a sample of a text, which would optimally represent the text as a whole, not only from the quantitative point of view, but from the qualitative one as well. The author sums up the results of a wide experimental research in the following way: (1) A random choice of pages does not represent well the text as a whole, representing only the chosen pages; the sampling was performed according to R. M. Frumkina's formula,<sup>2)</sup> comparing the distribution of parts of speech in the sample with their distribution in the frequency dictionary of Czech.<sup>3)</sup> (2) the optimum sample is that of a text ranging from the minimum of 3000 to the maximum of 5000 running words. Such a choice makes possible the comparison of various texts with regard to functional style, individual author's style, etc.

*J. Krámský* experimentally verifies the hypothesis according to which the frequency of nominal and verbal elements in the text can serve as stylistic characteristics. (As has been expected, the differentiating line runs namely between the colloquial and the scientific styles.)

*J. V. Bečka* deals with the important question of lexical stratification in texts from seven different fields. The proper purpose of his study is to establish whether a given word can be denoted as a term (characteristic for a given text), on the basis of its frequency. The decisive factor is not the overall frequency, but the number of the fields and the texts in which the relevant word has occurred.

To define the selective power of a word for informatic purposes *J. Helbich* uses two possible approaches: (1) objective, consisting in the determination of divergencies in the frequency distribution of the word in a given text, as compared with its distribution in a group of texts or in language as a whole; (2) subjective, consisting in the choice of words determined by specialists of a given field.

The sociolinguistically oriented contribution by *L. Klimeš* deals with the structure and development of slang from the quantitative point of view.

Several semantic and stylistic features of journalistic style are dealt with in the article by *J. Kraus*. His research is based on the Czech material of 50 000 words, including 5 532 adjectives. The author compares the occurrences of adjectives having frequencies of 1, 2, ..., 10, with theo-

<sup>1)</sup> PSML 1, Prague 1966, PSML 2, Prague 1967, PSML 3, Prague 1972, PSML 4, Prague 1972.

<sup>2)</sup> P. M. Фрумкина, Статистические методы изучения лексики, Moscow 1964.

<sup>3)</sup> J. Jelinek - J. V. Bečka - M. Těšitelová: Frekvence slov, slovních druhů a tvarů v českém jazyce, Prague 1961.

retical data obtained from Waring-Herdan's formula. The most frequent adjectives are then subjected to content analysis with the resulting data being compared with analogous data for Polish and French.

*L. Uhlířová* examines the relationship among the syntactic functions of the clause elements, their roles from the point of view of the functional sentence perspective, and their positions from the point of view of word-order, i.e. the range of problems usually not studied from the quantitative point of view. The article has proved that even in this untraditional discipline, the use of statistical method leads to suggestive conclusions concerning the most typical communicative functions of clause.

*M. Königová* deals in her methodologically-oriented article with the use of dichotomous algebra in the processing of syntactic data.

Phonological description of Czech includes the quantitative characteristics of Czech syllable, which is dealt with by *M. Ludvíková*. Using the correlation coefficient, she defines the relation between the frequency and the length of the syllable.

Thus, the articles in the first part of the new volume of PSML represent considerable methodological contributions. *M. Těšitelová* has solved one of the fundamental questions asked by every linguist dealing with lexical statistics: what part of a text is extensive enough to represent the text as a whole, and how to select this particular part. The results of her analysis are already finding their way into practical application. Not less important are the questions solved by *J. V. Bečka*. The results of his research as well as the results of *Helbich's* work will be appreciated both by linguists and specialists in information science. A rather special place among methodological contributions is occupied by the article of *M. Königová*. Whereas the other authors treat concrete language material, *Königová* explains the application of the method of dichotomous algebra, illustrating it by means of concrete examples. The other authors use the statistical methods with the aim of reaching stylistical (*Krámský*), syntactical (*Uhlířová*), or sociolinguistically oriented (*Klímeš, Kraus*) results. *M. Ludvíková* enriches the fonological description of Czech with the quantitative characteristics of the syllable. The character of most of the results tends to bring about wider conclusions, the results not being bound only with the concrete material from which they have been acquired. The adequacy between the linguistic material and the applied statistical apparatus is altogether good. Mathematical methods are concentrated (in the quantitative part) mainly on the absolute and relative frequencies and some other current characteristics: arithmetical mean, index of repetition, variation coefficient, entropy, and theoretical distribution formulas, e.g. those of Poisson and Waring-Herdan. Mathematicians might consider these methods rather too simple. Nevertheless, it should be noted that mathematical methods serve here only as a means of formulating conclusions of linguistic character, the use of a more complex mathematical apparatus not necessarily leading to more accomplished results in the solution of problems belonging to the sphere of linguistics.

In the second part of the volume *M. Novotný* studies algebraic properties of some relations on the basis of which configurations (in the sense of *Kulagina, Gladkij* and others) are defined; first of all some relations defined by languages (or subsets of monoids over some alphabet) are studied, such as domination and double domination.

*L. Nebeský* distinguishes two functions of a context-free grammar (specification of a set of strings and structural description of each string), giving some examples where it could be of advantage to discern two functions (with grammar working either in parallel, or in series), e.g. in complicated cases of concord, or with conjunction.<sup>4)</sup> Generative description is dealt with in other

<sup>4)</sup> In *Piřha's* proposal of rules for conjunction (see e.g. *P. Sgall, L. Nebeský, A. Goralčíková, E. Hajičová, Functional Approach to Syntax, New York 1969, p. 64*), where the derivational history is not identified with the structural description to such a degree as is the case with a grammar using P-markers, we can see an instance of such a differentiation of two functions of a grammar.

articles, as well: *J. Horecký* presents a proposal of generative description of Slovak, consisting of three parts, each of them being generative in its character: grammar, onomatology (description of naming units) and phonology (morphemic and phonemic structure). The articles by *E. Benešová* and *S. Machová* concern a different type of generative description, the so-called functional description the framework of which was proposed by P. Sgall (cf. op. cit. in Note 4). In this type of description, generative (recursive) character is proper only to the component generating semantic representations, with the other components having the form of pushdown store transducers gradually translating the semantic representation to the sound shape of the sentence. The input and output languages of these push-down stores are CF languages, but due to the connection of these devices in the sequence, the generated language is a context-sensitive language.<sup>5)</sup>

Semantics of verbal modality and its incorporation in the above mentioned framework is dealt with by *E. Benešová*, *S. Machová* describes the function of transducers in translating the adverbial of cause from the semantic (tectogrammatical) level to the phenogrammatical (syntactic) level.

An attempt at the machine generation of Czech ten-feet iambic verse by means of CFG rules and the rules of poetics, is made by *K. Pala*; the rules of poetic are understood as the number of feet admitted from the point of view of the given type of verse.

Formal properties of document retrieval languages (DR) are treated by *O. Sechser*. He shows how, by means of an operation of substitution of a DR language into another DR, a third DR is constructed, and how this operation can be used to construct hierarchies of DR languages.

The second part of the volume is mainly concerned with the employment of the apparatus of generative grammars, be it for the purpose of studying their formal properties, or of completion and enrichment of explicit linguistic description. Only the article of *M. Novotný* concerns the field of analytic models.

This brief characteristic of individual contributions in PSML implies that algebraic linguistics is understood here as a domain of linguistics using mathematical methods (and standpoints). It is important in this domain to find correct formal mechanism describing empirically revealed phenomena of language; the chosen mechanism may be trivial or not in mathematics itself, but here it serves first of all to formulate the empirical results in such a way that they might be understood, checked and applied without taking recourse of the reader's (or author's) intuition. Another aim is seen in such a formulation that the results are incorporated in a comparatively well established mathematical theory (if it is shown, for instance, that an empirically motivated type of description is equivalent with a context-free grammar, many characteristic features of the language generated have been found at once).

*Iva Nebeská, Jarmila Panevová, Praha*

---

<sup>5)</sup> Cf. now *M. Plátek's* study in *Prague Bulletin of Mathematical Linguistics*, 21, 1974.