

## REVIEWS

*The syntax of sentence and text: a festschrift for František Daneš.* By Světlá Čmejrková and František Štícha (eds). Pp. 398. Amsterdam-Philadelphia: John Benjamins, 1994.

Reviewed by Aleksander Szwedek, Nicholas Copernicus University, Toruń.

In the age of information explosion, it is difficult to keep track of all developments in any given field, particularly when the field is subject to paradigm pressure. That is one of the reasons (as well as political situation) why, for a long time, the achievements of the Prague School linguistics remained largely unknown (with notable exceptions) in the West. Although the book under review is a tribute to individual accomplishments of František Daneš, it is also a sign of general recognition and appreciation of the whole linguistic school.

As usual and expected in a festschrift volume, contributions are quite diversified as to the content. Some are of a more general nature, and some very specific, though not, of course, without general implications. The number and variety of contributions make it difficult for the reviewer to give an exhaustive, adequate and balanced account of all papers, and selection is not easy. I have decided to give a brief account of only some contributions of more general character.

The contributions are organised into four sections: Prague School Functionalism, Functional Sentence Perspective and Thematic Progression, Text and Discourse, and Grammar and Semantics, all topics reflecting the enormous range of Daneš' scholarship.

In the opening paper de Beaugrande discusses two approaches to language analysis: functionalism and formalism. He emphasises the largely underestimated proposals of the Prague school, among others the early awareness of the advantage of an interactive model of language over a modular one, as well as Daneš' distinction of three levels: the organisation of utterance, the semantic structure of sentence, and the grammatical structure of sentence.

One of the points de Beaugrande addresses is a methodological problem of linguistic analysis and language model. De Beaugrande writes that

a strict separation of levels would prevent an understanding of how any one level functions, even on its own terms. At best, a strict separation could be *simulated after the fact* once the data had been analysed from an interactive outlook. And this tactic might still entrain us in an arbitrary, often redundant reconstruction of parallel criteria and patterns for each level under different terms. (de Beaugrande 1994: 35).

It is, of course, trivial to say that any language entity is a multilevel construct, and additionally NOT a simple aggregate of levels and factors. Quite possibly it may be the "chicken-egg" problem, but I believe that the very identification, however intuitive and initially vague, of any single level or element is necessary to define its main function in which it necessarily interacts with other levels or elements, producing ultimately a complex structure.

Once levels or elements have been distinguished, they can not only be studied in their contextual interactions with other levels and elements, but also in their own terms, with other levels and elements kept as constant as possible. Thus separation should be looked upon as a convenient tool, perhaps an initial shortcut, not an end in itself.

Enkvist's "Centre and periphery, delicacy and fuzz" is a very central, delicate and non-fuzzy account of the major tendencies in modern linguistics, showing at the same time the relevance of earlier Prague School studies in the context of those new tendencies. In his extremely subtle study Enkvist addresses a problem similar to that raised by de Beaugrande, bringing to our attention once again the fact that quite a few "hot" ideas in contemporary linguistics were taken up much earlier by the Prague School. He quotes Daneš (1966: 11) as writing what is regarded as one of the basic principles today:

All such insoluble dilemmas and useless disputes can be clearly removed if one gives up the common

notion of strict compartmentalization and simply admits an obvious idea, viz. that the organization of linguistic elements in the "system of systems" has a different nature, a different structure, than is usually assumed: the classes (and sub-classes) of elements should not be regarded as "boxes" with clear-cut boundaries but as formations with a compact core (centre) and with a gradual transition into a diffuse periphery which, again, gradually passes (infiltrates) into the peripheral domain of the next category.

Our conception does not, of course, deny the existence of classes or categories, but at the same time it does not force us into unambiguous decisions in those cases where the decision has not been made by the language ...

Enkvist also emphasises the processual aspect of language, referring to his own studies (e.g. 1986), as well as to Humboldt's *energeia*, and Mac Whinney's (1989) Competition Theory. He also recalls Schaubert and Spolsky's (1986) ideas which he includes in what he calls "preference semantics". I think that in the context of the above mentioned proposals, due regard should be paid to de Beaugrande's various studies, perhaps particularly to de Beaugrande and Dressler (1980), where they underline the dynamic, probabilistic (preferences) and interactive nature of language, adding to the description the "problem solving theory" (decision-making in the Competition Model).

In his paper "Information Structure in Writing" Martin Davies discusses a very important problem of signalling information organization in written discourse concluding that

There thus seem to be grounds for believing that although there is no notation for intonation in the writing system of English, the grammar and lexis do make available sufficient information to enable readers - when reading - to discern the information structure the writer has in mind with sufficient precision to make it possible to know when they have got it right. And when they have made a mistake, they will KNOW they have, and will be able to say, "I'm sorry, I'll read that again." (Davis 1994: 88)

Wolfgang Dressler relates naturalness theory to de Beaugrande and Dressler's text model I referred to earlier. In particular he discusses some aspects of de Beaugrande and Dressler's standards of textuality in relation to FSP. Probably for lack of space, Dressler leaves undiscussed the most interesting part of the relation. He has advocated (e.g. 1990) general preference for binary relations (see also Szwedek 1985, 1990), which is manifested in the polar opposition between T proper and R proper. He does not however, take any stand on the fundamental, long standing assumption of FSP, i.e. CD and its graded character.

This is precisely the problem taken up in the next contribution by Firbas on Daněš' view of givenness as a graded phenomenon. Firbas bases his notion of "givenness" on the relation of retrievability. This raises the question of the length of the text stretch from which information remains retrievable, activated. Apart from the fact that the problem had been earlier addressed by Osgood (1971), it is worth noting that I see a significant change in the interpretation of FSP. Progression which is such an important part of CD seems to have taken on a new dimension. Instead of a number of degrees of contextual dependencies, Firbas presents a clearly prototype effect description; he admits that "in principle the opposition of context dependence and context independence is preserved" (cf. Szwedek 1985, 1990); however, it is possible that within the category of context dependence (givenness) there are more and less prototypical instances, giving the impression of gradual distinctions. That important modification is in perfect consonance with Dressler's (1990) "preference for binary relations" (however, cf. also Szwedek 1985, 1990). By the way, I think that the "binary relations" view is a very promising proposal; apart from the fact that we can find lots of binary relations (oppositions) (marked-unmarked, given-new, topic-comment, things-relations, figure-ground, foreground-background, things-relations, temporal-atemporal relations, etc.), all of them can be traced to the basic cognitive process described by Langacker (1986) as comparison: "Fundamental to cognitive processing and the structuring of experience is our ability to compare events and register any contrast or discrepancy between them. .... acts of comparison continually occur in all active cognitive domains, ..." (Langacker 1986:101). Langacker further writes that an act of comparison "has the general schematic form *S T*, where *S* can be called the *standard* of comparison and *T* the *target*." Thus any act of cognition is dichotomous in nature involving in its foundations a contrast between two entities. This would account for the role of "binary relations" in language. That dichotomy would then be subject to modification depending on the nature of both

the *standard* and the *target*, leaving the impression of gradation. This is what I advocated for many years with respect to information organization in the sentence (e.g. Szwedek 1976, 1985, 1990). Langacker also remarks that "the comparison of two events need not involve them as unanalyzed wholes; it may instead pertain only to certain facets of them or their manifestation in particular domains" (Langacker 1986: 104). He finally concludes that "This capacity for selection is also quite important for understanding semantic and grammatical structure".

Luelsdorf's contribution is a reinterpretation of thematic progression in terms of determinacy grammar, a new theory, judging by references. It makes quite difficult, technical reading, even more difficult by either lack of clarity or inconsistent use of symbols. For example, D is used for determinacy (throughout the paper), but also for directionality (p. 152); R is used for rheme, as well as resilience (p. 148), P for progression, as well as preceding (p. 152), B for position (p. 152), as well as binder (p. 153). Petőfi's contribution is in consonance with his earlier encyclopaedic approach to text analysis. His study is extremely interesting, full of technical details, showing all complexities of text constitution as basis for interdisciplinary studies, among others, because of its multimedial character. This new approach he proposes to call *Semiotic Textology* to refer to the most broadly comprehended text analysis and text description.

In conclusion, that interesting volume can be recommended to linguists of many persuasions, as it not only increases our knowledge of linguistic past, but also very much so about the linguistic present.

#### REFERENCES

- Beaugrande, Robert de – Wolfgang Dressler  
1980 *Introduction to text linguistics*. London and New York: Longman.
- Dressler, Wolfgang  
1990 "The cognitive perspective of "naturalist" linguistics models". *Cognitive Linguistics* 1: 75-98.
- Enkvist, Nils Erik  
1986 "Linearization, text type, and parameter weighting.", in: Jacob L. Mey (ed.).
- Langacker Ronald  
1986 *Foundations of Cognitive Grammar*. Vol. I. Stanford, California: Stanford University Press.
- Mey, Jacob L. (ed.)  
1986 *Language and discourse: Test and protest. A festschrift for Petr Sgall* (= Linguistic and Literary Studies in Eastern Europe 19). Amsterdam and Philadelphia: Benjamins.
- Osgood, Charles  
1971 "Where do sentences come from?", in: Danny Steinberg – Leon Jakobovits (eds.), 497-529.
- Steinberg, Danny – Leon Jakobovits (eds.)  
1971 *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*. Cambridge: Cambridge University Press.
- Fisiak, Jacek (ed.)  
1990 *Further Insights Into Contrastive Analysis*. Amsterdam: Benjamins.
- Pieper, Ursula – Gerhard Stickel (eds.)  
1985 *Studia Linguistica Diachronica et Synchronica*. Berlin: Mouton de Gruyter.
- Szwedek, Aleksander  
1976 *Word Order, Sentence Stress and Reference in English and Polish*. Edmonton, Canada: Linguistic Research.  
1985 "Sexpartition versus bipartition", in: Ursula Pieper and Gerhard Stickel (eds.), 801-809.  
1990 "What is topic? A contrastivist's view". in: Jacek Fisiak (ed.), 499-506.



*Quantitative linguistics*. By Marie Těšitelová. Pp. 253. Amsterdam & Philadelphia: John Benjamins. 1992.

Reviewed by Yuri Tambovtsev, Lwiw Lesotechnical, Ukraine.

This is the thirty-seventh volume of the series "Linguistic and Literary Studies in Eastern Europe (LLSEE)". The idea of this series is to give Western linguists solid information on recent developments in the countries of the former USSR (i.e., Russia, Belarus, Ukraine, Uzbekistan, Tajikistan, Kyrgyzstan, Turkmenistan, Azerbaijan, Armenia, Georgia, Moldova, Kazakhstan, Lithuania, Latvia, and Estonia), Poland, the former Czechoslovakia, Bulgaria, Romania, Hungary, Mongolia, Vietnam, and China (although the latter three are not European countries at all). This is a good idea since the majority of Western linguists are not well informed about recent developments in linguistics in the countries of the former communist block. The author of this book knows these developments well, although some important issues in the countries of the former USSR have escaped her attention. The book was translated from Czech, which may explain some clumsy expressions (e.g., p. 116).

The book is divided into seven main parts with more than a hundred subdivisions. It is well edited and easy to use. It has a "Name Index" and a "Subject Index". Although the number of references in the list of references is great (611), we will see later that this list misses some important entries. Nevertheless, the list of references is useful and interesting for all linguists with a potential research interest in quantitative linguistics, since it covers all the main bibliographical items. We can also say that, in its turn, the list can be considered as a list of researchers. Unfortunately, one finds very few names of scholars who have used qualitative methods in studying Turkic, Tungus-Manchurian, Finno-Ugric, Paleo-Asiatic, and Asian languages. There is insufficient space for me to name all of those missing, but the complete list can be found elsewhere (Tambovtsev 1984a, 1985b, 1986a, 1990b, 1991a).

In the first chapter, the author discusses the notion of quantitative linguistics, which she understands as the study of language (natural or artificial) using quantitative methods (statistics and probability calculus). She defines the scope and aim of quantitative linguistics as the application of quantitative methods to study the transition of qualitative changes into quantitative ones. Těšitelová is correct in stating that studying language by quantitative methods means not only counting the frequency of occurrence of language units of different levels, but also measuring them in the way done by J. Rehak in social science (p. 13). To my mind this is not a relevant explanation for a linguist. It should have been said more explicitly that quantitative methods reveal data which are hidden, revealing regularities and tendencies which exist in a language but are not immediately obvious. They can manifest themselves only through quantitative methods, e.g., the frequency of a certain phoneme. Thus, it leads us to the conclusion that both the instrument and the result of the investigation are important, though the author claims that from her point of view quantitative methods are nothing but merely an instrument. This may be right, but only to some extent. Whether one digs a pit with a spade or with an excavator, the result is the same, i.e., the pit. If we do not take into account the energy and the time spent, then it is possible to claim that the instrument does not matter. However, in my opinion, the situation here is rather different. It resembles more the one where the instrument produces different results. One can play Bach on a Russian balalaika (with three strings) or Vogul violin (with one string) and then claim that the result, i.e., music, is the same as the music produced by an organ. In my opinion the aim of quantitative linguistics is twofold: first to find the hidden linguistic characteristics and then to interpret the results; and second to show a linguist why unsophisticated statistical tools (a mere sum, mean, or percentage) are not enough and may obscure or falsify a further linguistic interpretation. I have argued elsewhere (Tambovtsev 1983, 1986b, 1988, 1992b) that many traditional linguists investigate linguistic units of different levels (phonemes, morphemes, word-forms, syntactic and semantic constructions, etc.) based on small unrepresentative sample sizes without reliable statistical tools and then claim a solid linguistic interpretation. One of the drawbacks of the book under review is that the author does not mention the third aim of quantitative linguistics, which is to establish a good reconstruction of the typology, i.e., branching structure, of a language family tree for a group of related languages, as done in the works of David Sankoff (1972) and Sheila Embleton (1983, 1985, 1986). These works are not analyzed in the book. I wonder if Těšitelová does not know of them, or if she considers it unimportant to discuss them, although in my opinion they show a new (proper and pioneering) approach, not only in quantitative but also general linguistics. I am sure that new investigations in this direction will lead to some extremely interesting and fruitful results. Thus, at least she should have mentioned it, even if she did not appraise it. Neither does she mention another fruitful and interesting approach whose

methods allow a linguist to establish exact typological distances between language families or to measure the compactness of language families. It has been possible to calculate the distances between the languages of the Finno-Ugric family (Tambovtsev 1983, 1991b, 1992a) and to compute the compactness of the Turkic, Tungus-Manchurian, Paleo-Asiatic, Finno-Ugric, and Indo-European language families and even such super-families as Ural-Altai and Nostratic (Tambovtsev 1990a). The author discusses typological statistics (Chapter 4, pp. 177-181), but devotes only several pages to it and examines it only with respect to language universals. Here she describes only the old and well-known results of J. Greenberg, V. Krupa, V. Skalicka, J. Kramsky, A.L. Kroeber, and C.D. Chrétien, but does not tackle or even mention the works of more modern investigators (e.g., L.G. Zubkova, A.S. Gerd, M. Rimmel, V.M. Andrijushchenko, K.B. Bektaev, Y.A. Tambovtsev, A.V. Zubov, G.J. Martynenko) who have introduced some recent valuable results in quantitative linguistics.

Speaking in general about new books on quantitative linguistics, the typical linguist will be asking by now whether or not this book is frightening and just how much previous knowledge of mathematics is really necessary as a prerequisite. As far as this book is concerned, from the point of view of mathematics the typical linguist would probably understand it, while for a specialist in mathematical or quantitative linguistics it may seem a bit too simple. At the same time this may be an advantage, because it may serve as an understandable beginning textbook for a student who is interested in quantitative linguistics.

The second part of the book is dedicated to research methods. The author discusses in detail the units of population in morphological, lexical, and other domains of quantitative linguistics. In lexical statistics, instead of defining a word as a unit, Těšitelová cuts short all the definitions of word by accepting a rather formal definition of it, i.e., as a graphic unit – a letter or a group of letters between two spaces. This may be all right for well-known and well-established languages such as English, Russian, French, Czech, Hungarian, Polish, etc., but it can hardly be applied to Mansi (Vogul), Jug (Ket), etc., or even Japanese, where some particles are considered by some linguists as part of the word while by others as separate words. One also encounters more or less the same problem in German. The author does not show whether she considers such German verbs as *angehen*, *aufbauen*, and *einschlagen* as one word or two. How many words should be counted in the following sentences? *Was geht dich das an? Der Plan baut darauf auf, daß... Die Neuigkeit schlug wie eine Bombe ein.*

It is good that Těšitelová considers another important issue which is too often ignored by the typical linguist: sampling of material (pp. 24-31). The reader may find a lot of valuable information on how to sample the material in order to obtain reliable research results. It is a pity that typical linguists do not pay much attention to this crucial point, and then find themselves in a position where their research is statistically unreliable, and therefore unreliable from the point of view of the linguistic interpretation. Reading linguistic investigations outside the field of quantitative linguistics, one can vividly see that the typical linguist is apt to take into account anything (style, individual style, all peculiarities of slang and dialect, etc.), but not the sample size or the manner of sampling (systematic, random, or cluster sampling) which may influence the results (Tambovtsev 1986b, 1991a). Thus the book is a reliable source of information for any linguist on how to select the material for research in a scientifically appropriate way (pp. 32-46). It should be mentioned that in the natural sciences the principle of correct sampling is always observed. The use of statistics as it begins at the end of Chapter 2, which deals with the First, Second, and Third Zipf's Laws, mean, dispersion, frequency distribution, correlation and some concepts of information theory (pp. 50-66). All in all, one can state that it is a good description of the usual statistical methods used in quantitative linguistics.

The third chapter of this book depicts the main areas of quantitative linguistics. In connection with these areas, I should mention that when I began to study the literature on quantitative linguistics as a post-graduate student in 1973, I noticed that the first works in it began to be published about a century ago, through the great impetus it had had in the 40s and 50s, and in the 60s it had undergone great development and achieved a peak, especially in the field of phonostatistics. However, then the wide, fast stream of publications began to cease. It now lacked constructive ideas. The upper crust was removed, the first results received – but many linguists did not bother to go deeper. Analyzing the works of those linguists who actively studied phonostatistical problems (especially Soviet colleagues), I could vividly see that their interests had shifted back to their previous tasks. I guessed that this area of linguistics was not fashionable any more. Now in the countries of the former USSR, only the scholars of the "Statistics of Speech" Group continue with research on Germanic, Slavic, Romance, Finno-Ugric, Turkic, Tungus-Manchurian, Paleo-Asiatic and some other languages by quantitative methods, but their



number decreases with every year. In my opinion, the same trend can be observed in the West. It can be seen in the chapter under consideration, which deals with 1) lexical; 2) grammatical (morphological and syntactic); and 30 semantic statistics. Lexical statistics is represented by the different sorts of frequency dictionaries, frequency lists and concordances. However, one finds no discussion of the frequency dictionaries of the Finno-Ugric, Baltic, and Asian languages, nor are they in the general list of frequency dictionaries (pp. 69-70), although the author suddenly discusses them later (pp. 98-100).

Speaking about one of the main problems of lexical statistics, i.e., the richness of vocabulary, Těšitelová dwells upon the formula of P. Guirad (p. 76), who introduced the notion of concentration of vocabulary (pp. 77-78). His formula was later developed by J. Mistrik (p. 79). Těšitelová also introduces her own formula (p. 80). She claims that when studying the richness of vocabulary, one should take into account 1) the repetition of words in a text; 2) the dispersion of the vocabulary; and 3) the concentration of the vocabulary (pp. 80-82). A student of linguistics can find lots of valuable information on Slavic languages: the lexical statistics of Czech (pp. 84-85), Slovak (p. 86), Russian and Ukrainian (p. 87), Polish (p. 88), and over Slavic languages (p. 89). The description of Slavic lexical statistics is more complete than in the other works of this type published in the West. The author gives a detailed outline of lexical statistics concerning the Germanic languages (German, p. 90; English, p. 92; other Germanic languages, p. 94). The discussion of publications on lexical statistics on the Romance languages include French (p. 95), Spanish (p. 96), Rumanian (p. 97), and Italian (p. 97). Těšitelová also describes the results of investigations in lexical statistics which are usually unknown to the general linguist, e.g., Baltic (Latvian, p. 98), Finno-Ugric (Estonian, Hungarian, and Finnish, pp. 98-99). The only Asian language represented in this chapter is Chinese (p. 100, though the author should have mentioned the solid investigations in Kazakh. One can find the list of publications on Kazakh, Uzbek, and some other Turkic languages elsewhere (cf. Tambovtsev 1986a, 1987, 1988). The same drawback must be mentioned with regard to the grammatical, semantic, and other domains of quantitative statistics. Unfortunately, the limited space of a review does not allow detailed analysis of morphological and syntactic statistics or of semantic statistics, although one can notice that the author has narrowed her description mainly to Czech (pp. 110-114), Slovak (p. 115), Russian and the other Slavic languages (p. 115), while of the other languages she only considers English (p. 117), German (pp. 117-118), Latin, Rumanian, and Latvian (p. 118), devoting to each only a pair of lines. In the field of syntactic statistics she also deals only with Czech (pp. 127-129), Slovak (p. 130), Russian (p. 130), and Polish (p. 131). In my opinion, she should have mentioned at least several important works (e.g., Peshchak et al. 1979), in which a group of Ukrainian linguists presented the main statistical schemes of the Ukrainian word. A more detailed account of this and other works of Ukrainian colleagues and other linguists of the former USSR can be found elsewhere (Tambovtsev 1991a). The author should have known the works in statistical syntactic analysis because they are few. Computers are seldom used in this field, while the majority of articles and books is devoted to lexical statistics. This is why it is quite strange that such a well-informed scholar as Těšitelová would forget to mention one of the most interesting works in the field of syntactic statistics, Martynenko (1983), where the feature domain was constructed on the basis of the linear and hierarchical construction of the sentence. His multi-dimensional classification space was based on the Russian prose of 86 writers (for details and exact references of items referred to below, see Tambovtsev 1991a: 149-151). Some of the titles of the articles in Ukrainian and Russian may have misled Těšitelová, who did not analyze them, although they depict problems of grammatical statistics (e.g., V.I. Perebejnos, ed., *The automatization of the analysis of scientific text*, where the authors concentrate their attention on the statistical analysis of the letter chains and some syntactic algorithmic rules which are also studied by these authors in "Linguistic problems of editing" and by T.A. Grjaznuhina in "Analysis of the prepositional links in scientific Text". In addition, Těšitelová outlines the object of semantic statistic (p. 135), method of research (p. 135), method of research (p. 135), unit of population (p. 136), an selection of methods and material in it (pp. 138-140). Dwelling on the publications on semantic statistics, the author emphasizes the contribution of the linguists of the former USSR, such as B.A. Plotnikov, V.A. Moskovich, R.M. Frumkina, and J. Tuldava. However, she pays much more attention to the investigation of Czech, mentioning only one work in German, and not mentioning the recent achievements in English at all.

Chapter 4 deals with the other domains of quantitative linguistics which include 1) phonological; 2) graphemic; 3) stylistic; 4) typoloical statistics; 5) statistics concerning the development of language and 6) word-formation statistics. In my opinion, it is this part of the book which deals with the most promising fields of quantitative linguistics. Beginning with phonological statistics, Těšitelová correctly

remarks that phonology belongs to the domains of quantitative linguistics where the application of statistical methods has a long tradition, although in many languages it is still an open question as to what the phonemes are or how many phonemes should be in this or that phonemic inventory. However, my own experience of counting frequencies of speech sounds in languages where the exact phonemic inventory is disputable nevertheless shows that frequency analysis, especially frequency analysis of sounds in certain positions in a word and their combinability, allows a researcher to differentiate between the actual phonemes and phonemic variants. Thus the investigation of the frequencies of the members of the sound chain in Mansi (Vogul) allowed me to find out the exact phonemic inventory (Tambovtsev 1977, 1979, 1980, 1981) and to verify the phonemic inventory in Jug (Ket) (Tambovtsev & Werner 1979). It is good that once again the author draws the student's attention to the correct sampling of the material in phonological statistics. She is correct to state that there is a belief among traditional linguists that for reliable phonemic counts large samples are not necessary. It may be because of the solid works where linguists took samples of 1000; e.g., Greenberg counted small samples of 1000 each for Hausa, Klamath, Coos, Yurok, Chiricahua, and Maidu (Greenberg 1966). However, since M. Königova investigated the size of corpus for determining the frequency of phonemes, it is not possible to derive any reliable linguistic results from small samples. Königova (1966) showed that the necessary corpus size for the most frequent phonemes (i.e., only the first three phonemes of the ordered series) is about 7000-8000 phonemes, while the least frequent phonemes require a corpus of not less than 150,000 phonemes. Těšitelová lists this article (p.147) along with the other articles. In fact, the sizes of the samples really affect the linguistic conclusions (Tambovtsev 1985a). In the field of phonological statistics, Těšitelová provides a full account of the works of Czech scholars: V. Mathesius, J. Vachek, B. Trnka, and others. The author also gives some account of the works of scholars in the other Slavic languages, although she does not describe some solid investigations in Russian (e.g., Bondarko, Zinder & Shtern 1977; Jolkina & Judina 1964). In the Ukrainian part, she does mention the first and pioneering work by Perebejnos (1964). Gridneva has greatly influenced Ukrainian colleagues by her work (e.g., 1966). Maybe one of the main drawbacks of the book under review is the very short list of works which are discussed in chapter 1.3.4, dedicated to the phonological statistics of "other languages". Here Těšitelová speaks only about English, German, French, Spanish, and Hungarian, leaving many languages uncovered. Let us name just a few. Segal (1972) is a good book on the history and modern methods of phonostatistics whose main part is devoted to the thorough investigation of Polish, one of the best book on phonostatistics even now, 20 years after its publication. There are several other works that are not in the limelight and are almost never cited. Guinashvili (1965) has a sample of 30,503 phonemes of Persian, with results not less interesting than those in the other book on Persian phonostatistics, Moinfar (1973), also unmentioned here. As mentioned earlier, no work in the phonostatistics of Mansi, Hanty, Veps, Karelian, Tatar (Baraba), Oroch, Komi-Zyrian, etc. (cf. Tambovtsev 1984a, 1985bc, 1986a, 1988, 1990b), nor the reliable work of Van den Broeke et al. (e.g., 1986) on Dutch is mentioned. Some very well-known books are not discussed at all, for no apparent reason (e.g., Brainerd 1974). European phonostatistics is usually much better represented than Asian, so it is worth mentioning the following: Kissen (1964), with a large sample of 40,894 phonemes in Uzbek; Atamuradov (1966), 35,000 phonemes in Turkmen, another Turkic language; and Elizarenkova (1974), who counted Aryan (i.e., Old Indic) phonemes in 10-12th century B.C. texts. Especially important is Vertogradova (1967), which studies the phonostatistical structure of the prakrits (five Middle Indic dialects), and was missed by Těšitelová, even though it follows and develops the ideas of the Western linguists Harary & Paper (1957). The work would have been a richer source if these, and other, works had not been omitted.

Later, Těšitelová considers stylistic statistics (pp. 160-177), defining three branches: 1) selection and use of linguistic means in communication, in a text; 2) rhythmical layout of verse, which the author proposes calling quantitative versology or statistical versology; and 3) statistical characteristics of language when dealing with so-called disputed authorship. One cannot help agreeing with her that the important thing is to choose correct characteristics on a certain language level (or several levels), then to use them correctly, and then to evaluate the results correctly. So it is very important to choose a "good" unit of population and sampling (pp. 160-161). She again speaks of the richness of vocabulary and the index of repetition of words (a matter she had already discussed earlier) – it would have been advisable to deal with it all in one place, here giving just a cross – reference to the earlier discussion. Her treatment of disputed authorship is only based on the works of Alvar Ellegård and P. Vashák, although she mentions works by G.U. Yule, C.S. Brinegar, A.Q. Morton, and some others. She should have described in greater detail the work of V.J. Batov and J.A. Sorokin, who used 8 characteristics



and obtained convincing results (p. 176). Her selected publications on stylistic statistics give a balanced view (pp. 165-174). It was a good idea to devote a separate part to typological statistics, although it seems to me that she understands the object of typological statistics too narrowly, since she believes that only typological statistics quantifies so-called language universals (p. 117). In my opinion, typological statistics should also include different indices on which it is possible to construct different taxonomies or classifications. Těšitelová discusses only old works in this field, e.g., J. Greenberg (1960), V. Skalička (1967), J. Kramský (1976), T. Milewski (1962), M.I. Lekomtseva (1963), V.I. Perebejnos (1970), A. Avram (1964), E.A. Afendras (1970), etc. (pp. 178-181; see therein for exact references). In my works on typological statistics, I showed the functioning of the consonant-vowel ratio in some 80 languages around the world (Tambovtsev 1985d). The distances between the Finno-Ugric languages were defined elsewhere (Tambovtsev 1983). The linguistic typological distances between Japanese, Oroch, Jug (Ket), Jakut, Kazakh, Uzbek, Turkmen, Mansi, Hanty, Selkup, and other languages were measured on the basis of functioning of certain consonantal groups in a large sample. Genetically close languages tend to show similar consonantal functioning in the defined groups (Tambovtsev 1988). The typological distances between Hakas and four other Turkic languages were found (Tambovtsev 1991b), and the new notion of compactness was introduced and measured in Finno-Ugric, Samoyedic, Uralic, Turkic, Tungus-Manchurian, Altaic, Ural-Altaic, and their groups, subgroups, and branches (Tambovtsev 1990a). The compactness of the language family is measured as the sum of distances between languages within a language family. Phonemic functioning of speech chains in different languages may show similarity in their construction, since the investigation of genetic relatedness deals with the dialects or supposed dialects of a language. It turned out that the Tihvin and Ludikov dialects of Karelian are only a bit closer to each other than are Mansi (Vogul) and Hungarian (Tambovtsev 1984b). Describing the application of quantitative methods in dialectology, Těšitelová never mentions either this work, nor the numerous works of Hans Goebel in this direction, nor of others (e.g., Sheila Embleton), although in general more effort is needed here (Tambovtsev 1992b).

The next section of the book (pp. 181-188) is devoted to the glottochronological method and selected publications on statistics concerning the development of language or languages. Glottochronology is not popular any more even with many scholars who were very active in this field, e.g., G.A. Klimov, who wrote a solid work on the glottochronology of the Transcaucasian languages (dealt with by Těšitelová) but who now does not consider works on glottochronology to be of any value. Most of the interesting works in this field were written in the 1960s, and then the interest of linguists in this theory stopped. However, the well-known formula of Morris Swadesh (p. 182) can be used a bit differently and then yield valuable results. It is a pity that nobody pursued the direction of investigation proposed by M.V. Arapov & M.M. Herts, which Těšitelová speaks about later (p. 186); she does not say anything about how this method and some other methods were later applied by Arapov (1988). There are also more recent works in the field, with new methods, by David Sankoff (e.g., 1972) and Sheila Embleton (e.g., 1983, 1985, 1986).

The application of the results of quantitative statistics is believed to be of use to stenographers, typographers, language teachers, psychologists, and the decoders of coded messages (pp. 190-199), and some others. Těšitelová mentions the decipherment of Maya texts by a group of mathematicians from the Institute of Mathematics of the Siberian Branch of the Russian Academy and a group of linguists from Novosibirsk University, but does not give exact references to their articles.

In writing about quantitative linguistics and computers (Chapter 6, pp. 200-202), Těšitelová does not tell us her thoughts on why the number of works in the field began to decrease when computers came in.

That is why what she tells us in the last chapter (pp. 203-208) about prospects for quantitative linguistics does not sound convincing, although one may agree that this field of linguistics now needs a team of experts trained both in linguistics and mathematics. In my opinion, it is the lack of new ideas which hinders its development. As I have shown above, there are some promising areas in this field; they should develop rapidly, and then, as is usual with fashions, linguistic fashion will return to the field of quantitative linguistics. I would recommend revising and updating this book for a second edition, and devoting more pages to the most promising areas.

## REFERENCES

- Alekseev, P.M.  
1975 *Statisticheskaja leksikografija (tipologija, sostavlenie i primenenie chastotnyh slovaraj)*. Leningrad: LGPI.
- Goncharenko, V.V.  
1971 "Chastotnyj slovarj anglijskih tekstov po poluprovodnikom", in: *Statistika rechi i avtomaticheskij analiz teksta*. Leningrad: Nauka: 179-190.
- Brainerd, B.  
1974 *Weighing Evidence in Language and Literature: A Statistical Approach*. Toronto and Buffalo: University of Toronto Press.
- Jones, D.  
1917 *Everyman's English Pronouncing Dictionary*. London: J.M. Dent & Sons.
- O'Connor, J.D.  
1974 *Phonetics*. Harmondsworth: Penguin Books.
- Roberts, H.  
1965 *A Statistical Linguistic Analysis of American English*. The Hague: Mouton.
- Tambovtsev, Y.  
1979 "Raspredelenie glasnych fonem v mansijskoj poezii", *Soviet Finno-Ugric Studies* XV, 3: 164-167.
- Tambovtsev, Y.  
1984 "Phoneme Frequency and Closeness Quotient: Establishing Relationship Degrees by Phonostatistics", *Ural-Altaic Yearbook* 56: 103-119.
- Tambovtsev, Y.  
1992 "Phonostatistical Study of the Vepsian Language", *The Bulletin of the Phonetic Society of Japan* 201: 9-14.
- Tambovtsev, Y.  
1994 Review of Těšitelová. 1992. *Word* 45: 365-377.
- Těšitelová, M.  
1992 *Quantitative Linguistics*. Amsterdam & Philadelphia: John Benjamins.