Popescu, I.-I., Best, K.-H., Altmann, G.: *Unified modeling of length in language* (= Studies in Quantitative Linguistics 16). Lüdenscheid: RAM-Verlag 2014, 123 pages. ISBN: 978-3-942303-26-2

The analysis of length of various linguistic units is perhaps one of the oldest activities in quantitative linguistics. Unfortunately, people historically considered length in quite different units, some of which were not very appropriate. For example, word length in terms of number of letters (which is not applicable to Chinese or Old Egyptian), sentence length in terms of number of words yielding incommensurable results for example for Hawaiian and Inuktitut, or yet worse, sentence length in terms of number of phonemes yielding a fraction. The principle holds: measure the length of a unit in terms of its immediate constituents, e.g. words in syllables or morphemes; sentences in terms of clauses; clauses in terms of phrases, etc. Never count a length in terms of letters because letters are part of the secondary language. The authors of *Unified modeling of length in language* show that counting the immediate constituents of linguistic units is not only a dogma of quantitative linguistics, but it can also be validated by means of the results of mathematical modelling.

The second disagreeable circumstance of many length analyses of linguistic units is partly the lack of a unique theoretical background of the modelling. One can often observe an inductive way of searching for an appropriate model, and in quantitative linguistics the deductive approach of the "Unified Theory" developed by Gejza Wimmer and Gabriel Altmann is already at our disposal (Wimmer/Altmann 2005). The basic idea behind it is to model the interrelation, or the dependency, of two linguistic variables by means of discrete probability models and continuous functions. One dependent variable y is influenced by one independent variable x, both interrelated by a relative rate of change dy/y. Based on this approach one tries to fit distributions or functions at least belonging to the same family; later on one can unify them to a general form of a basic difference or differential equation (technically for this step of data modelling one can use common software, such as *R*, *Nlreg*, whereas for discrete probability distributions Altmann-Fitter 3.0 is appropriate and powerful software[1] for modelling linguistic data).

The basic question discussed in the reviewed monograph is whether all lengths of linguistic units could be reduced to one single equation which can be presented as a function or a probability distribution. The authors succeeded and proposed the so-called Zipf-Alekseev function, defined as

$$y = ax^{b + c \ln x},$$

[1] Cf. http://www.ram-verlag.biz/altmann-fitter/ (accessed 07-06-2015) for further details.

which in its differential equation form contains the state of language, the force of the speaker and that of the equilibrating hearer (community). In all studied cases taken from the existing literature they obtained a satisfactory fit. They considered syllable length, morph length, word length, length of compounds, length of rhythmic units, verse length, sentence length and speech act length. For all results they presented both the tables of measurements and graphical presentations. The determination coefficient R^2 is in all cases overwhelming. The languages analysed were German, Quechua, Chinese, Inuktitut, Estonian, Finnish, Hungarian, Vogul/Mansi, Cheremis/Mari, Erzya-Mordvin, Latvian, Gaelic, Welsh, Dutch, English, Faroese, Gothic, Icelandic, Low German, Swedish, Greek, French, Italian, Portuguese, Romanian, Spanish, Belorussian, Bulgarian, Czech, Low Sorbian, Old Church Slavonic, Polish, Russian, Slovak, Slovenian, Ukrainian, Maori, Marquesan, Arabic, Old Hebrew, Turkish and Uzbek. This was not a choice of positively tested cases but depended on the availability of the data. It is to be hoped that further languages will be tested at each level in order to bring further corroborations. However, the study of language levels and the length distributions of linguistics entities is not finished, but can be understood as a stimulus for the further statistical modelling and the ongoing open problem of the interpretation of the parameters. In this respect the authors propose a level-dependent interpretation of the yielded values – an initial suggestion which without any doubt has to be pursued.

The fact that the authors use a continuous function instead of a discrete distribution does not change anything: models are our views and means for capturing and formally processing the observed reality. A unified formula is better than a great number of models yielding different explanations and in any case the possibility of the transformation of continuous models to discrete distributions is given (cf. Mačutek/Altmann 2007).

Needless to say, even if many languages and units have been scrutinised, one needs further data concerning all levels in order to consider the model a language law. The reviewed book can be recommended to all experts interested in the field of quantitative linguistics, especially those focusing on length problems of linguistics, and can serve as a starting point for further in-depth studies in this field.

References:

Mačutek, Ján; Altmann, Gabriel (2007): Discrete and Continuous Modelling in Quantitative Linguistics. In: *Journal of Quantitative Linguistics* 14 (1), 81–94.

Wimmer, Gejza; Altmann, Gabriel (2005): Unified derivation of some linguistic laws. In: Reinhard Köhler, Gabriel Altmann und Rajmund G. Piotrowski (Hg.): Quantitative Linguistik. Quantitative Linguistics. Ein internationales Handbuch. An International Handbook. Berlin, New York: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft, 27), 791–801.