Mathematical models of linguistic laws

This project deals with the mathematical modeling of two language laws. According to the first of them, the Menzerath-Altmann law, longer language units consist of shorter parts (ie, for example, longer words contain shorter syllables on average).

A number of publications on this topic have appeared over the last few years, but so far no general model has been presented for this law that also takes into account the properties of linguistic units. During the solution of the project, a monograph will be written, the first author of which will be the applicant.

The second law is a model for the ranked frequencies of phonemes and graphemes. The only model that displays a reasonably good fit is the negative hypergeometric distribution. However, there are problems with the interpretation of its parameters. As a part of the project solution, a linguistically reasonable interpretation of the parameters of this distribution will be suggested. Moreover, the classic chi-square goodness-of-fit test is unusable, as it rejects practically all null hypotheses for large samples. An empirical decision rule will be presented that allows the same approach to evaluating the fit between model and data for both large and small samples.