Evolution of the Temperature Parameter in Texts

Andrij Rovenchak[a] and Solomija Buk[b]

[a]Department for Theoretical Physics,
[b]Department for General Linguistics, Ivan Franko National University of Lviv, Ukraine

In recent papers [1,2], we suggested a new set of parameters to be used in the analysis of texts. The parameters are defined by mapping word rank–frequency distributions onto the Bose-distribution within the grand-canonical approach. The respective physical analogues are the power of the excitation spectrum α and the temperature T. The analogue of the fugacity z is determined from the number of words occurring only once (so-called *hapax legomena*). The values of α typically fall in the domain between 1 and 2. The calculated "temperature" values behave so that the relation $\tau = \ln T / \ln N$ (where N is the number of words) does not vary significantly as the length of text increases. The values of *alpha* and τ, however, show the relation to the type of language grammar. In this work, we analyze in detail the evolution of the τ and α parameters in texts of several kinds, namely: translations of the Gospel of John and *The Little Prince*, a novel by Antoine de Saint-Exupéry, into several dozens of languages from different language families, as well as several works of the long-prose fiction by Ivan Franko, a famous Ukrainian writer from the turn of the 20th century.

[1] A. Rovenchak and S. Buk, Physica A **390**, 1326 (2011).
[2] A. Rovenchak and S. Buk, J. Phys. Stud. **15**, 1005 (2011).