

Diffusion Entropy Analysis of Long-range Correlation In The Turkish Language Using Non-corpus Parametrization

Gökhan Şahin^{*}, Orhan Özgür Aybar, Avadis Hacınlıyan

Department of Information Systems and Technologies, Yeditepe University,
Kayışdağı, İstanbul, Turkey

^{*} Electronic Address: sahin@yeditepe.edu.tr

The existence of long-range correlation in (twelve) natural languages by means of detrended fluctuation analysis had been reported by Hacınlıyan [1] et al where each language shows two distinct scaling regimes. This work extends this line of research to demonstrate that the historical development of a language such as Turkish can easily be traced out using the proposed corpus independent parametrization.

The first issue that one faces in an attempt to analyze a natural language as a time series is the mapping of texts onto a time series. The usual choice in is using a time series derived form a corpus (see [2, 3, 4]). Our proposed variable in this work is inspired by DNA random walks and is derived by assigning values to the letters constituting a given word. The resultant time series for texts in the Turkish language from different time periods are analyzed via Diffusion Entropy analysis (DEA) ([5]).

The texts analyzed in this work show two different regimes under DEA, the short range regime is virtually parallel within errors in all cases, while the long range correlations for the texts belonging to the time period are markedly different from the other texts of different time period. These results imply that the time evolution in a language can be detected by means of DEA based on the proposed parametrization.

-
- [1] Sahin , G , M. Erentürk and A. Hacınlıyan, "Detrended fluctuation analysis in natural languages using non-corpus parametrization", *Chaos, Solitons and Fractals*, Article in Press.
 - [2] Montemurro MA, Pury PA. "Long-range fractal correlations in literary corpora." (2002). Available from cond-mat/0201139.
 - [3] Bhan,J, S. Kim, J. Kim, Y. Kwon, "Long-range correlations in Korean literary corpora", *Chaos, Solitons and Fractals*, **29**, 69- 81, (2006).
 - [4] P. Allegrini, P. Grigolinib and L. Palatella, "Intermittency and scale-free networks: a dynamical model for human language complexity", *Chaos, Solitons and Fractals*, **20**, 95-105, (2004).
 - [5] N. Scafetta, V. Latora and P. Grigolini, "Lévy scaling: The diffusion entropy analysis applied to DNA sequences", *Phys. Rev. E*, **66**, 031906, 2002.