A Modern Greek Readability Tool: Development of Evaluation Methods

Rania Voskaki & George Mikros

National and Kapodistrian University of Athens

Keywords: readability tool, corpora, annotation, evaluation methods

The present research reports on an online Modern Greek Readability tool (MOGRead[1]), developed by the Centre for the Greek Language. It constitutes part of an ongoing postdoctoral research project that aims to evaluate MOGRead in terms of effectiveness and efficiency, in order to propose novel methods that will enhance the tool's accuracy.

MOGRead was developed in order to meet the needs of teachers of Modern Greek as a second/foreign language (L2). Provided that the existing Modern Greek corpora are few and rarely updated, the need of an effective readability tool is imperative, especially for less used and less taught languages, such as Modern Greek.

Under this scope, the objective is to provide reliable results for any text as it concerns the level of adequacy (A1 to C2) according to Common European Framework for languages. In order to evaluate MOGRead' reliability, we have created two distinct corpora of plain texts: (a) a representative set of texts, according to their language level A1 to C2, available online on the Portal for teaching Modern Greek as L2[2], and (b) a verification corpus that we have annotated and classified to the equivalent language level. In order to tag the verification corpus we opted for the TreeTagger tool (Schmid, 1994). As for the analysis of both corpora by quantitative methods, we used QUITA (Kubát et al., 2014).

The applied evaluation methods will permit the development of novel methods for the readability analysis of texts, taking into account the readability that differs among readers (DuBay, 2004) and especially among learners of a second/foreign language; in our study Modern Greek learners' readability. Consequently, the objective of this study is to underline the findings that we came up with after the implementation of evaluation methods on both corpora, to present the outcomes, and to set the basis for the procedure that has to be followed in terms of larger-scale reliable results.

References

DuBay, W.H. 2004. *The Principles of Readability.* Costa Mesa, California: Impact Information.

Kubát, M., Matlach, V., Čech, R. 2014. *QUITA - Quantitative Index text Analyzer.* Lüdensheid: RAM.

Schmid, H. 1994. "Probabilistic Part-of-Speech Tagging Using Decision Trees", in *Proceedings of International Conference on New Methods in Language Processing,* Manchester, UK.

[1] http://www.greek-language.gr/certification/readability/index.html.
[2] http://www.greek-language.gr/certification/dbs/teachers/index.html.