Co-referential, thematic analysis, and network analysis of the poem A szőlőműves by Milán Füst and its English translation[*]

István Károly Boda[a], **Erzsébet Tóth**[b]

[a]Department of Mathematics and Informatics, Debrecen Reformed Theological University
boda.istvan@drhe.hu

[b]Department of Data Science and Visualization, University of Debrecen
toth.erzsebet@inf.unideb.hu

Abstract

The background of our study is a 3D virtual library project which started in 2013, as part of cognitive infocommunications (CogInfoCom) direction of research. The content of the virtual library includes verbal and multimedia materials (mainly literary texts in English, and in some cases in Hungarian) which, among other things, can be excellently used for English language learning purposes. In the recent years, we gradually put the emphasis on bilingual language learning materials, adding Hungarian or English translations to selected texts of the virtual library. In the current implementation of the 3D virtual library we extensively use the 3D features of the MaxWhere Seminar System. Nevertheless, from the beginning of the project we have been creating the hypertext representation of the library materials using the standard web technology (i.e. HTML/CSS/JavaScript etc.).

As regads the selection of the content of the library, we have always considered it our mission that the eternal cultural values which classical literary works can convey to the present-day culture would also be available for the users of the 3D virtual library, and especially for the young generations including the members of

---

the generations CE. Keeping in mind their wants and needs as well, we created a bilingual (English-Hungarian) language learning material developed for English language learners at an advanced level. We attempted to organize the bilingual material to form a more or less scale-free network of interconnected nodes so as to provide an efficient and user-friendly learning environment.

As a further development of our 3D virtual library project we would like to provide access to *full texts*, both in Hungarian and in English, for the potential users of the bilingual language learning material. Because the mission of the virtual library project includes the accumulation and transmission of cultural values expressed in classical literary works, we decided to select short poetic texts with high emotional content which may involve attentive and intensive reading and, in turn, effective learning. Because of the high complexity of natural language texts, and especially poetic texts, it is crucial that the texts should be processed and prepared efficiently to satisfy the needs of the language learners. Therefore, in addition to the basic text processing and organizing technologies which we have used so far, we also applied linguistic and textological tools for the comprehensive and intricate processing of poetic texts.

In this paper we will analyse and process a selected Hungarian poem and its English translation (Füst Milán: *A szőlőműves*; Milán Füst: *The Vine-Dresser*, translated by István Tóthfalusi) using the methods and notation of *co-reference analysis* the complex notation and terminology of which have been invented, developed and applied to the analysis of Hungarian literary texts by János S. Petőfi. With reference to our bilingual language learning material, one of the main advantages of co-reference analysis is that it can serve as a kind of metalanguage which is independent of natural languages. That characteristic, as we will show, makes co-reference analysis particularly useful in computer-based text processing.

The central aim of the application of co-reference analysis for the polyglott investigation of selected poetic texts is to identify and organize the Hungarian and English keywords of the texts in order to create entries for the bilingual language learning material and explore the thematic structure of the analysed texts. As regards the methodology and use of co-reference analysis, we rely on the results of our previous studies mainly published in the series of Officina Textologica. In the first we investigated the theme-rheme relationships of the poem *'A szőlőműves'* by Milán Füst; in the second we carried out a parallel analysis of the Hungarian and English version of the poem *'Köd előttem, köd utánam...'*, also by Milán Füst. Both studies was based on the methodology of co-reference analysis.

First, we present the parallel co-referential analysis of the poems *'A szőlőműves'* and *'The Vine-dresser'*, and based on this, we formally describe the theme-rheme (or topic-comment) relationships between the keywords of the texts, display their thematic structure, **create a bilingual hypertext representation of both texts**, and examine, from a network theoretical perspective, the characteristics of the network of hypertext links that represent the combined thematic structure of the two analysed texts.

Because of the complexity of natural language texts, it is essential to represent

the linguistic knowledge as well as the background knowledge which are both necessary to understand and interpret the analysed texts. Taking a formal approach, we use several types of dictionaries as widely accepted sources of linguistic knowledge on the one hand; and we seek and select additional texts as relevant sources of background knowledge on the other hand. During the analysis of the original texts (i.e. the selected poems by Milán Füst) we add commentaries to the communication units that make up the text sentences of the analysed texts in order to formally represent the linguistic and background knowledge. Then we create entries for all keywords that occur in any of the text sentences in the role of theme (or topic). The resulting entries will contain the corresponding text sentences, their communication units, and the commentaries that complement them. The entries are considered as network nodes, and the connections between them are established by the keywords that occur in either the communication units or the commentaries (or both of them).

In this way, we can build the hypertext (HTML) representation of the texts based on the system of entries and connections. Due to the interactivity of the hypertext implemented as a web page, this representation reflects the hypertextual interpretation of the texts, and makes it accessible to the potential users. In a technical view, the accessibility of the hypertext links of the web page from a JavaScript program enables the examination of the global properties of the thematic structure of the texts as a network, especially its scale-free characteristic. During the creation of the hypertext representation as a web page, we pay great attention to using the sema data structure as that of the 3D virtual library, so that the web page obtained as a final result of the analysis can be fully integrated into, and thus it will be an inherent part of the virtual library [1–5].

References

[1] A.-L. Barabási: *Linked. The New Science of Networks*, Cambridge, MA: Perseus Publishing, 2002.

[2] P. Baranyi, Á. csapó, G. Sallai: *Cognitive Infocommunications (CogInfoCom)*, Berlin, Heidelberg: Springer, 2015, DOI: https://doi.org/10.1007/978-3-319-19608-4.

[3] I. K. Boda, J. Porkoláb: *Co-reference Analysis and the Structure of Natural Language Texts*, in: Szöveg az egész világ. Petőfi Sándor János 70. születésnapjára, Budapest: Tinta K., 2002, pp. 81–100.

[4] I. K. Boda, J. Porkoláb: *Füst Milán Köd előttem, köd utánam... című versének és angol fordításának korreferenciális elemzése*, Officina Textologica 21 (2020), pp. 31–42.

[5] I. K. Boda, E. Tóth: *English language learning by visualizing the literary content of a knowledge base in the three-dimensional space*, Annales Mathematicae et Informaticae 53 (2021), pp. 45–59, DOI: 10.33039/ami.2021.04.003.