

Lexicographic Support of Computer-assisted Language-learning Authenticity

Danko Šipka (Arizona State University) and Jakub Zgoliński (Adam Mickiewicz University, Poznań)

In an attempt to provide more authentic and autonomous instruction of less commonly taught languages Bosnian/Croatian/Serbian (BCS), Polish, and Russian text taggers (<http://www.asusilc.net/cgi-bin/newtepajgu.pl>) are being developed as a part of the project titled Learner-centered Task-based Language Instruction (<http://www.asusilc.net/lctli>) The taggers allow the user to paste in a text, copied from an on-line newspaper or acquired in another manner, and have it tagged with English glosses and equipped with the option of displaying the inflection of each word form in the text.

Authenticity (understood in all its complex and multifaceted semantic network) and learners' autonomy belongs to the most important task of the present-day language instruction, as persuasively discussed in Mishan (2004). However, attempting to address this imperative leads to the following dilemma. On the one hand, students with limited command of the vocabulary cannot be expected to process "raw" authentic texts as constant dictionary lookup would be overly time consuming and frustrating. On the other hand, instructors do not command sufficient financial and temporal resources to manually gloss these texts with English equivalents. The aforementioned taggers are meant to resolve this dilemma by offering an authenticity-fostering courseware.

The major challenge currently addressed within the Slavic text taggers project is that of appending and amending the knowledge bases. Two major techniques are being deployed in appending and amending the knowledge bases. First, the databases are filtered using available monolingual and bilingual lexical lists or dictionaries and the missing entries are added. Second, the tagger itself is used as an important tool of testing the comprehensiveness and accurateness of the database. In order to enhance this process, the authors have been developing a knowledge base authoring tool.

Lexicographic System Lextras is a side project of the team standing behind the taggers and is gradually replacing old tools which were initially used to maintain the knowledge bases. Since Lextras uses native XML database for storage and TEI specifications for encoding it is suitable for any kind of lexical data. Apart from securing the data storage and its consistency the system has a wide range of features which increase productivity of definers and maintainers. The authors will present how Lextras is being used in development and management of the taggers project.

The present paper summarizes achievements of this project hitherto, identifies its major problem areas, and outlines its envisaged development hence. The presentation is accompanied by a demonstration of the software under discussion.

References

Mishan, Freda. *Designing Authenticity into Language Learning Materials*. Bristol, GBR: Intellect Books, 2004