

Toma Tasovac

(Department of Comparative Literature
Princeton University)

TRANSPOETIKA PROJECT: DIGITAL TECHNOLOGIES AND DISTRIBUTED LEARNING IN FOREIGN LANGUAGE STUDY

Abstract. The internet and new technologies offer numerous methodological innovations for the practice of teaching foreign languages, yet teaching Serbian as a foreign language does not seem to have been affected by the contemporary trends in hypertextuality, interactive learning and use of audio-visual materials. Annotated multimedia editions of the most important works by Serbian and Yugoslav writers would greatly improve and enrich the learning process of Serbian as a foreign language, while possibly attracting a larger number of students to the study of South Slavic literatures and cultures. The *Transpoetika Project* is an academic initiative for publishing annotated, fully-glossed study editions of literary works in the Serbian language through the use of non-proprietary formats, interoperable platforms and web standards such as XML, TEI and Unicode. At the core of the *Transpoetika Project* lies a collaborative, web-based Serbian-English dictionary which encourages the students of Serbian to become producers of knowledge in a global learning network. Unlike most online dictionaries today, the *Transpoetika* dictionary is also a dictionary service for distributed learning: a user-configurable RSS system allows dictionary users and editors to monitor recent changes in the dictionary and to contribute to its growth by requesting and receiving translations of new words via RSS feeds.

Learning Objective

To provide a set of integrated linguistic tools to assist an intermediate or advanced foreign student in reading, comprehending and discussing authentic literary texts in the Serbian language.

The learning objective itself needs some justification. First of all, why are we focusing on intermediate and advanced students? The students who have covered the basics of the foreign language grammar and who can use the language in everyday spoken situations with a certain degree of proficiency often cannot break out of what is known as the “intermediate plateau”: a level of linguistic proficiency, which may satisfy the student’s basic communication requirements, but which excludes the receptive awareness and active usage of natural collocations, the ability to manipulate different linguistic registers, and exploit nuanced, metaphoric speech [1]. The experience of foreign language teaching shows that a significant number of students never manage to increase their spoken and written fluency beyond the intermediate level, despite taking intermediate and advanced courses [2]. Reading authentic, unabridged, and “difficult” literary texts can be extremely frustrating if the student lacks the vocabulary or sufficient knowledge of complex syntactic structures and cultural references, but, at the same time, reading simplified, abridged versions of literary works can often be an insult to the student’s intelligence. The *Transpoetika Project* is firmly founded on the belief in the

primacy of meaning in language learning and the lexico-semantic approach which stresses the importance of units larger than the single word (polywords, fixed and free collocations, institutionalized expressions etc.), reflects the native speaker's reliance on pre-fabricated chunks of language and suggests that the dictionary should not simply be a reference tool, but rather a learning tool [3]. The unmodified language of literature can and should be used as a preparation for authentic language experiences outside the class because literature, among other things, encourages interaction and intellectual discussion; reading literature in the original can only increase the student's language awareness (standard vs. substandard language, style, figures of speech); and reading literature in a foreign language is a highly rewarding intellectual experience, which should be encouraged in its own right [4]. The *Transpoetika Project* is a technological platform for the delivery of literary texts together with a set of linguistic tools that will help the student take the full advantage of the literary text at hand.

Background and Need

The internet and new technologies offer numerous methodological innovations for the practice of teaching foreign languages, yet teaching Serbian as a foreign language does not seem to have been affected by the contemporary trends in hypertextuality, interactive learning and use of audio-visual materials. The author of this project, who taught Serbian and Croatian languages at Harvard and Princeton, believes that the extremely limited choice of materials and study aids available for Serbian and/or Croatian may be one of the reasons why Slavic Departments at many US universities have decreasing or very poor enrollments in their South Slavic courses. Annotated multimedia editions of the most important works by Serbian and Yugoslav writers would greatly improve and enrich the learning process of Serbian as a foreign language, while possibly attracting a larger number of students to the study of South Slavic literatures and cultures.

The *Transpoetika Project* has been envisioned as a multidisciplinary project, based on highest academic and technological standards. Our goal is to provide superbly annotated literary works, and a basis for a whole series of digital editions of Serbian and Yugoslav authors.

Speaking in more general terms, the *Transpoetika Project* is an initiative for using digital technologies in foreign language teaching and learning, which comes as a response to the insufficient and/or inadequate presence of the so-called "small" languages on the internet. The goal of the project is to facilitate the use of digital technologies in the realm of language learning and offer technological and methodological solutions that will contribute to the development of linguistic and cultural diversity on the internet.

Modules

The architecture of the *Transpoetika Project* consists of four distinct, interoperable modules, each of which can exist on its own and serve its own purpose. The integration of the four modules provides a platform which facilitates the reading and encourages the discussion of the literary works written in the Serbian language:

- **TransText:** the basic textual unit, encoded in XML with a TEI DTD, specifically encoded in view of the needs of a foreign student, with a set of XSL and CSS

stylesheets which transform and format the XML document into valid XHTML for web delivery, with embedded links toward the other three modules

- **TransDictionary:** a Serbian–English dictionary for the students of Serbian as a foreign language, based on the highly structured MySQL database, which is used for depositing the linguistic analysis and translations of the words and phrases from the TransText module; also a collaborative dictionary tool which enables the students of Serbian to participate in the making of the dictionary in a global learning network
- **TransMedia:** a collection of foreign language audio and video materials, saved in the QuickTime format for easy internet delivery; a media component of the TransText and TransDictionary modules (audio recordings of the texts, pronunciation of headwords and examples in the dictionary etc.); can also be used as a basis for the development of independent listening comprehension exercises.
- **TransForum:** a web-based forum for the discussion of the Serbian language and literature, which can be linked to from the other three modules.

User Interface

The rationale behind the Transpoetika UI is that the text on the screen should be as clean as possible: the tools should remain invisible as long as the user does not activate them. Even when activated, the tools should in no way distract the reader from the text itself. In practice, this means that we are using a complex cascading style sheet to control the visibility of elements on the web page.

A TransText is delivered to the reader as an UTF-8 encoded, valid CSS-styled XHTML document, which is viewable in all modern browsers across the computer platforms. Clicking on any word in the text activates a PHP script, which queries the TransDictionary and displays the dictionary entry for the given word on the same page, next to the text. Moving the computer mouse over the white space before each sentence in the text makes visible the small inline “play” sign below the mouse: clicking on the “play” sign starts the audio recording of the given sentence. Moving the mouse over the indented space of each paragraph, makes the inline paragraph sign visible (¶). Clicking on the paragraph sign plays the audio recording of the entire paragraph. In both cases, while the sentence or paragraph audio file is playing, the control bar appears next to the text so that the student can stop, rewind and forward the current sound file. When the mouse is not hovering above the beginning of a sentence or a paragraph, no signs or control bars are displayed, so that there are no visual distractions during the silent reading. We are currently working on implementing a similarly unobtrusive way for the user to select a sentence, paragraph or entire chapter to start a new or join the existing TransForum discussion on the selection in question. The TransDictionary and TransForum have their own user interfaces, and can be accessed independently of the TransText.

Modular Integration

The TransText is prepared for publication as a valid XML document [5], encoded in full compliance with the Guidelines of the Text Encoding Initiative (TEI) [6]. In addition to textual segmentation tags, which are useful for structuring the electronic version of the text, particular attention is paid to the analytic apparatus, which the foreign student will

find useful while reading a foreign text. We implemented a consistent way of marking up different levels of analytic annotation, be it linguistic (morphology, syntax, grammar and lexicon), literary (style, meaning, allusion, alliteration), or cultural (geographical, historical, political and social references).

All of the annotated information can be divided into two groups, one being strictly contextual, the other, global. Contextual information is relevant only to the text under discussion; global information, on the other hand, is the kind of information that could be reused in a different context. Only contextual information is actually encoded in the text itself; global information is deposited in the TransDictionary database in such a way that a unique identifier can link the text to the external resource. Rather than defining the meaning of each word in the text, which would become impractical and repetitive in a longer work or across different texts, we are only encoding the dictionary form of the given word, and that only if the dictionary form itself differs from the inflected form in the text. When transforming the XML document into XHTML, the XSLT style sheet embeds a PHP script to query the word from the database. As a result, the user can click on the given word and retrieve its corresponding entry from the dictionary. Since each sentence and each paragraph are marked-up with corresponding tags (<s>, <p>) and unique attribute ID's (which count the position of the given sentence or paragraph in the text as a whole), the same numbering scheme is used when saving recorded audio files of the given text. In this case, too, XSLT is responsible for embedding audio links with each sentence or paragraph.

The TransDictionary is the first Serbian–English dictionary specifically designed for the students of Serbian as a foreign language. It provides full conjugations for the verbs as well as nominal and adjectival declensions, monolingual word definitions, English translations of the words, collocations, examples, quotations and, optionally, links to the available pronunciation sound files [7]. Taking advantage of the excellent language-specific abilities of the Google search, the dictionary retrieves a host of usage examples of the given word from the internet, in both the Cyrillic and Latin alphabets, encouraging the student to widen the scope of his lexical knowledge in other, sometimes unpredictable, contexts.

TransDictionary is also a collaborative dictionary, which means that students can contribute to its growth and make it into a useful, general purpose Serbian–English dictionary, independently of it being an accompanying tool for the TransText module. Unlike the existing online collaborative dictionaries [8], the TransDictionary is not fully open-content. The disadvantages of “extreme” open content policies in the compilation of a dictionary are obvious: anybody with an internet connection can not only add and alter, but also delete the content of the dictionary; quality control and accuracy are extremely difficult to manage, hence the need for long-winded disclaimers. [9] The TransDictionary is a collaborative dictionary for registered users (contributor, editor, senior editor) who have different access privileges, which can guarantee streamlined accuracy control. Having different levels of user access is, however, not enough for managing the flow of information and changes in the dictionary. This is the reason why we have implemented a system of user-configurable RSS feeds. Using an RSS news aggregator, both the users and editors are constantly updated on the new words, which have been entered into the dictionary, or any other changes in the dictionary. Although the value of *rich site summary* or *really simple syndication* in online learning has been noticed before [10], I am not aware of its use in other dictionary projects.

Furthermore, as the number of dictionary contributors grows, RSS could be used as a full-fledged dictionary service: a user enters a word into the dictionary for which there is no translation, and subscribes to recent updates. The editors and/or contributors, who are subscribed to new words, receive the new word in their RSS feed and use the supplied link to translate it in the dictionary database. Subsequently, the user who is subscribed to recent updates receives the translation in his or her RSS feed.

Conclusion

The *Transpoetika Project* is designed in such a way that it provides learning opportunities outside the classroom and encourages not only the communication between students, but also the student's active participation in extending the scope of the Serbian–English online dictionary. Using digital technologies and respecting highest academic standards, this project attempts to bridge the technological gap which separates “major” from “minor” languages, revitalize the methodology and practice of “lesser taught” languages, and contribute to the popularization of Serbian and Yugoslav literature abroad.

References

- [1] Felicity O'Dell, *Teaching advanced learners*, <http://uk.cambridge.org/elt/teachers/youngadult/articles.htm>
 - [2] Michael Lewis (ed.), *Teaching Collocation*, Hove, 2000
 - [3] Michael Lewis, *The Lexical Approach*, Hove 1993; Michael Lewis, *Implementing the Lexical Approach: Putting Theory into Practice*, Hove 1997
 - [4] Lazar, Gillian, *Literature and Language Teaching*, Cambridge 1993; Widdowson, H. *Stylistics and the Teaching of Literature*, Longman 1975.
 - [5] <http://www.w3.org/XML/>
 - [6] <http://www.tei-c.org/>
 - [7] The dictionary is based on the Webtionary, an open-source MySQL/PHP dictionary developed by Danilo Segan for the dictionary of Serbian computer terminology. See <http://www.prevod.org/recnik/> and <http://kvota.net/webtionary/>
 - [8] Alternative Dictionaries, see <http://www.notam02.no/~hcholm/altlang/>; or Wiktionary, an Wikipedia offspring, “a multilingual and international dictionary... [which] tries to cover every word from all known languages.” <http://en.wiktionary.org>
 - [9] http://en.wiktionary.org/wiki/Wiktionary:General_disclaimer
 - [10] Eva Kaplan-Leirson, RSS: A Learning Technology, http://www.learningcircuits.org/2004/may2004/0405_trends.htm; Stephen Downes, Content Syndication and Online Learning, <http://www.downes.ca/cgi-bin/website/view.cgi?dbs=Article&key=972053825>
- ttasovac@princeton.edu