



KSKS – Corpus of Serbian as a Foreign Language

• overview and call for collaboration •

The Center for Serbian as a Foreign Language at the Faculty of Philology in Belgrade, in collaboration with the Department of General Linguistics, has initiated the construction of the first Corpus of Serbian as a Foreign Language (KSKS - *Korpus srpskog kao stranog jezika*).

The corpus and its purpose

The Corpus of Serbian as a Foreign Language will include texts written by learners of Serbian during exams, in class, or as part of homework. In the first stage of corpus construction, learner texts will be scanned, transcribed and saved as textual documents. Available background information on the individual learners will also be noted (their age and gender, first language, proficiency level and length of exposure to Serbian, other foreign languages known); the texts will be anonymized, i.e. the learners' names will not be included. The second stage will focus on the addition of linguistic information such as part-of-speech tags and lemmas, as well as the annotation of errors made by the learners. In the final stage, the corpus will be made available online, via a user interface.

The purpose of the Corpus of Serbian as a Foreign Language is to make a large number of annotated learner texts available to lecturers and researchers interested in studying Serbian as a foreign language, or in using authentic learner material in their teaching.

Possible corpus uses

The Corpus of Serbian as a Foreign Language will greatly automate the search for words and word sequences in learner data. It will also enable the texts to be searched by part of speech or lemma, and it will allow for automatic extraction of word, lemma and part of speech frequencies. Learner characteristics such as native language and proficiency level in Serbian will be implemented as optional search filters. The corpus will thus significantly facilitate error analysis, with possible uses ranging from theoretical research to the actual teaching of Serbian as a foreign language and the creation of pedagogical materials.

Call for collaboration

In order to enlarge the corpus, we invite all those involved in the teaching of Serbian as a foreign language, in Serbia or abroad, to contribute by sending us any relevant material already in their possession, or by starting to collect new material. We need learner texts coupled with the following data (if available): (1) the learner's native language, (2) proficiency level in Serbian, (3) length of exposure to Serbian, (4) other foreign languages known, (5) gender, and (6) age.

The corpus is currently written-only, but if sufficient recordings (of classes, oral exams, or other relevant activities) turn out to be available, an oral component will also be added.

Depending on your interests, you can also collaborate on other activities related to corpus construction, such as transcription or error annotation.

For further information and details on how you can collaborate on data collection and/or corpus construction, please contact:

Dr Maja Miličević

Department of General Linguistics

Faculty of Philology, University of Belgrade

m.milicevic@fil.bg.ac.rs