

Krasimira Aleksova, Laska Laskova, Yordanka Velkova
Sofia University “St. Kliment Ohridski”
Sofia, Bulgaria

CORPUS OF SPOKEN STUDENT ACADEMIC SPEECH

(Summary)

This paper presents some basic steps in creating a corpus of spoken student academic speech collected during the field work of two research projects developed in 2008-2010. The corpus is based on transcriptions of audio and video recordings connected to an oral academic communication of students in Bulgarian philology. The corpus contains two types of oral texts: 1) paper / presentation – oral monologic texts dealing with theoretical problems (some recordings are accompanied by discussions), 2) oral exams proceeding as a discussion – oral monologic-dialogic text.

We use EXMARaLDA software system to transcribe the recordings (<http://www.exmaralda.org/>) and a version of GAT Minimal Transcription System adapted for Bulgarian language (<http://paul.igl.uni-freiburg.de/gat-to/>). Corpus creation process involves: 1) video recording process and collecting metadata activity (information about the participants and the situation), as well as some compulsory steps (each participant signs a declaration of consent); 2) extracting audio from video using Pinnacle Studio software; 3) anonymization and normalization of the recordings with Audacity software; 4) transcribing the events from the audio/video files with EXMARaLDA transcription tool Partitur Editor; 5) video and transcription synchronization. Our further goal is to create a Coma corpus with the EXMARaLDA tool for administering corpus metadata – Corpus Manager (CoMa).

The paper also presents some core notation symbols for describing verbal and non-verbal events in an EXMARaLDA transcription; the basic interface elements of Partitur Editor; a detailed description of the types and categories assigned to a tier; some transcription output format.

Keywords: corpus, student academic speech, oral text, transcription, transcription convention, EXMARaLDA, Partitur Editor, Coma, GAT, metadata, verbal and non-verbal information, Partitur tier

✉ Krasimira Aleksova
krasimiraaleksova@gmail.com
✉ Laska Laskova
laskova@slav.uni-sofia.bg
✉ Yordanka Velkova
yordankavelkova@gmail.com