

Yovka Tisheva, Marina Dzhonova
Sofia University "St. Kliment Ohridski"
Sofia, Bulgaria

**CORPUS OF SPOKEN BULGARIAN – SPECIFICITY
AND STRUCTURE**

(Summary)

The paper presents the features of a corpus constructed entirely of data on spoken Bulgarian. The strategy for building the corpus is motivated by the features of spoken communication and the aim is to preserve the characteristics of spoken language when transcribed as a text file. The transcription and annotation systems provide a clear and accessible representation of language phenomena. The pragmatic and socio-cultural information accompanying the texts allows the corpus to be applied also in the field of humanitarian studies.

Keywords: corpus, spoken communication, spoken language, transcription, annotation, dialog

✉ Yovka Tisheva
tisheva@slav.uni-sofia.bg
✉ Marina Dzhonova
djonova@slav.uni-sofia.bg

Published: 30 September 2011