

Svetla Koeva, Diana Blagoeva, Sia Kolkovska
Institute for Bulgarian Language, Bulgarian Academy of Sciences
Sofia, Bulgaria

***BULGARIAN NATIONAL CORPUS PROJECT:
ACHIEVEMENTS AND PERSPECTIVES***

(Summary)

The paper presents the Bulgarian National Corpus project (BulNC) – a large-scale, representative, online available corpus of Bulgarian. The BulNC is also a monolingual general corpus, fully morpho-syntactically (and partially semantically) annotated, and manually provided with detailed meta-data descriptions. Presently, the Bulgarian National corpus consists of about 320 000 000 graphical words and includes more than 10 000 samples. The corpus structure and the accepted criteria for representativeness and well-balancing are briefly presented. The query language for advance search of collocations and concordances is demonstrated with some examples. The *Bulgarian National Corpus* enables a number of applications in various areas of linguistics: among others lexicography; computational linguistics; theoretical studies of specific linguistic phenomena; for extracting example sentences for teaching Bulgarian.

Keywords: corpus linguistics, electronic corpora, Bulgarian language

✉ Svetla Koeva
svetla@dcl.bas.bg
✉ Diana Blagoeva
diablag@mail.bg
✉ Sia Kolkovska
sia_btb@yahoo.com

Published: 30 September 2011