

## СЪВРЕМЕННО СЪСТОЯНИЕ НА БЪЛГАРСКО-АНГЛИЙСКИЯ ПАРАЛЕЛЕН КОРПУС СЪС СЪОТНЕСЕНИ ГЛАГОЛНИ ФОРМИ

**ТОДОР ЛАЗАРОВ**

ИНСТИТУТ ЗА БЪЛГАРСКИ ЕЗИК „ПРОФ. ЛЮБОМИР АНДРЕЙЧИН“ ПРИ БАН  
*tdlazarov@gmail.com*

Текстът представя съвременното състояние на Българско-английския паралелен корпус със съотнесени глаголни форми. Представени са целите, за които е създаден корпусът, характеристиките на корпуса в контекста на съвременната теория и практика. Дава се описание на лингвистичната анотация и структурните особености на корпуса. Представени са особеностите на текстовите типове и първични езикови данни, извлечени от корпуса.

*Ключови думи:* корпус, езиков ресурс, темпорална система, глаголни форми

## CURRENT STATE OF THE BULGARIAN-ENGLISH PARALLEL CORPUS WITH ALIGNED VERB FORMS

**TODOR LAZAROV**

INSTITUTE FOR BULGARIAN LANGUAGE, BULGARIAN ACADEMY OF SCIENCES  
*tdlazarov@gmail.com*

This paper deals with the current state of the Bulgarian-English parallel corpus with aligned verb forms. First, I present the objectives for the construction of the corpus and its characteristics in the context of contemporary theory and practice. A description of the linguistic annotation and structural features of the corpus is given. I also present the features of the text types and the primary language data extracted from the corpus.

*Keywords:* corpus, linguistic resource, temporal system, verb forms

## ЛИТЕРАТУРА

- Коева 2014: *Коева, Св.* Българският национален корпус в контекста на световната теория и практика – В: *Езикови ресурси и технологии за български език*. София, АИ „Проф. Марин Дринов“, с. 29-53.
- Куцаров 2007: *Куцаров, Ив.* Теоретична граматика на българския език. Морфология. Пловдив, УИ „Паисий Хилендарски“.
- Biber 1993: *Biber, D.* Representativeness in Corpus Design. – *Literary and Linguistic Computing*, vol. 8, issue 4, 1 January 1993, pp. 243-257. Buchholz & Marsi

- 2006: *Buchholz, S. and Marsi, E.* CoNLL-X Shared Task on Multilingual Dependency Parsing. – In: *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X '06)*. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 149-164. EAGLES 1996: *Expert Advisory Group for Language Engineering Standarts Preliminary Recommendations on Corpus Typology*. EAG-TCWG-CTYP/P. Pisa, Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale. <<http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html>> Koeva & Genov 2011: *Koeva, S., A. Genov.* Bulgarian Language Processing Chain. – In: *Proceedings of Integration of Multilingual Resources and Tools in Web Applications*. Workshop in conjunction with GSCL 2011, 26 September 2011, University of Hamburg. Koeva et al. 2010: *Koeva Sv., D. Blagoeva, S. Kolkovska.* Bulgarian National Corpus Project. – In: *Proceedings of LREC-2010*, Valletta, ELRA, 3678-3684. Koeva et al. 2012: *Koeva, S., B. Rizov, E. Tarpomanova, Tsv. Dimitrova, R. Dekova, Iv. Stoyanova, S. Leseva, Hr. Kukova, and A. Genov.* Application of Clause Alignment for Statistical Machine Translation. – In: *Proceedings of SSST-6, Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Jeju, Republic of Korea, 12 July 2012, The Association for Computational Linguistics, ACL 2012 / SIGMT / SIGLEX Workshop, 2012, pp. 102-110.
- Leech 1991: *Leech, G.* The State of the Art in Corpus Linguistics. – In: *Aijmer, K., B. Altenberg* (eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London, Longman, pp. 8-29. Schmid 1995: *Schmid H.* Improvements in Part-of-Speech Tagging with an Application to German. – In: *Armstrong S., Church K., Isabelle P., Manzi S., Tzoukermann E., Yarowsky D.* (eds.) *Natural Language Processing Using Very Large Corpora. Text, Speech and Language Technology*, vol 11. Dordrecht, Springer, pp. 13-22.
- Yimam et al. 2014: *Yimam, S., Castilho, E., Gurevych, I.* Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno. – In: *Proceedings of ACL-2014*, demo session.

✉ Тодор Лазаров

Секция по компютърна лингвистика  
Институт за български език „Проф. Л. Андрейчин“ при БАН  
бул. „Шипченски проход“ 52, бл. 17, 1113 София, България