



OPINION

by a member of the scientific jury: Assoc. Prof. Dr. Ekaterina Tarpomanova, Sofia
University St. Kliment Ohridski

for dissertation for obtaining the scientific and educational degree *doctor*

Author of the dissertation: Todor Lazarov, Department of Computational Linguistics,
Institute of Bulgarian Language Prof. Lubomir Andreychin, Bulgarian Academy of Sciences

Dissertation topic: Bulgarian-English grammatical parallels with respect to machine
translation

Todor Lazarov graduated in *Bulgarian Studies* at the Sofia University St. Kliment Ohridski and obtained a master degree in *Computer Linguistics. Internet Technologies in the Humanities* at the same university. He continued his studies in the field of computational linguistics as a PhD student at the Department of Computational Linguistics of IBE, BAS, with a supervisor prof. Svetla Koeva, while also working in a translation agency. After successfully passing a preliminary defense, he is now in procedure of a public defence presenting all the necessary scientific and administrative documents.

The dissertation consists of 203 pages and represents a linguistic approach to one of the main areas of the computational linguistics, machine translation. Without any doubt, this is a relevant issue with respect to the development of world science towards computer processing of big data, but on the other hand, this is a focused study, as it presents a specific issue – the verb tense in a pair of languages. The text consists of an introduction, four chapters, a conclusion, a bibliography and an appendix with a hyperlink.

In the introductory part the PhD candidate presents in a concise form the relevance and the motivation of the research, its goals and tasks, its object and focus, the methodology which is applied and a short plan of the dissertation. In addition to fully meeting the standards for the design of an introduction to a dissertation, this part is characterized by precise wording without any unnecessary deviations.

The second chapter is a theoretical comparison of the verb category tense in Bulgarian and English. The representation of the verb as part of the speech is clear and concretized compared to the preliminary version. The bibliographic awareness of the PhD candidate in presenting different concepts for the tenses in both languages is impressive. I find the decision of Todor Lazarov to use the system of description of the tenses of H. Reichenbach, which was applied in Bulgarian linguistics by R. Nitsolova, to be appropriate. Verbal tenses are considered separately for each language, correlated with their interaction with the categories person, number and gender. The chapter ends with a comparative analysis of the temporal systems of the two languages and the relation between tense and aspect. Through this approach, the many-sided representation of tense has been achieved both in terms of its categorial status and situated in the verb systems of Bulgarian and English.

The third chapter contains a description of the Bulgarian-English Parallel Corpus with Aligned Verb Forms, on the basis of which the statistical analysis and the conclusions from the work are made. The PhD candidate has well evaluated and motivated the use of the corpus approach in machine translation and its specific application in this work. In my opinion, this chapter is the most problematic one in the whole dissertation, as it leaves the wrong impression of compiling a corpus, while in fact a ready-made corpus is used, to which a small percentage of texts from the Internet and sentences from R. Nitsolova's *Bulgarian Grammar* have been added. The addition of sentences from *Bulgarian Grammar* in the corpus, in which temporal forms with low frequency in Bulgarian with purposeful translation into English, so as to present the most accurate correspondence of tenses, also does not reflect the overall picture of the translation equivalents in the pair of languages. However, I should note that this chapter has been significantly improved, the redundant descriptions of the resources of the Department of Computational Linguistics have been shortened, and the logical structure has been presented (BNC > Bulgarian-English Parallel Corpus > Bulgarian-English Sentence- and Clause-Aligned Corpus), which leads to the choice of corpus used in this work.

The fourth chapter presents the most important part of the study, namely the grammatical parallels in the temporal systems in Bulgarian and English according to the data extracted from the corpus. The statistics presented in the tables give a very good orientation for the use of tenses and translation correspondences between the two languages. Todor Lazarov has approached the

analysis of quantitative information in great detail and depth and has made interesting conclusions about the temporal correspondences and differences in Bulgarian and English.

The fifth chapter presents the practical application of the corpus approach used. Although briefly presented, the experiment of integrating the resources and grammatical information used, as well as the evaluation by users show the high value of the dissertation.

The conclusion contains a summary of the results achieved in the dissertation and its contributions, which are correctly presented. The bibliography includes important studies by Bulgarian and foreign authors on the topic (probably it can not be completely exhaustive on this type of topic).

The abstract correctly reflects the content of the work, and the publications on the topic of the dissertation exceed the required number, two of them are in editions indexed in the Web of Science.

My general impression of the dissertation is positive. Todor Lazarov demonstrates in-depth knowledge of computational linguistics, masters and easily uses the terminology in this scientific field, his style is at once precise, scientific and readable. The dissertation is original and contributing, the goals and tasks set by the PhD candidate have been fulfilled, not to be neglected, and the fact that he has taken into account a large part of the notes made on the preliminary defense and the text has been completed.

Based on what has been stated in this opinion, I strongly support the doctor degree to be conferred on Todor Lazarov.

April 8, 2020

Sofia

Assoc. Prof. Ekaterina Tarpomanova