

O P I N I O N

by **Assoc. Prof. Boryana Koleva Bratanova, PhD**

Department of English and American Studies, Faculty of Modern Languages

St. Cyril and St. Methodius University of Veliko Tarnovo

professional field 2.1. Philology. Modern English – Grammar and Semantics

of a doctoral thesis for awarding the scientific degree Doctor of Science

in professional field 2.1. Philology. The Theory and Practice of Translation,

New Bulgarian University, School of Graduate Studies

Department of Foreign Languages and Cultures

to Senior Assistant Professor Velislava Raykova Stoykova, PhD

Doctor of Science Thesis: Machine Translation Approaches to the Bulgarian Language

1. Significance of the research subject matter with reference to science and applied studies

The submitted Doctor of Science thesis presents various approaches in the area of statistical machine translation with reference to their applicability to Bulgarian translation. The topic merits thorough research since the development of language technology is directed mainly towards languages with greater number of speakers, being most developed for processing the global English language. Dr Stoykova aims at exploring machine translation approaches for the Bulgarian language, which is not among the world languages with high number of speakers, as a result of which Bulgarian language processing technology is developed mainly in Bulgaria by native speakers of Bulgarian. Being a Slavic language with rich morphology, Bulgarian is favourable for natural language processing. In this respect Dr Velislava Stoykova's Doctor of Science thesis demonstrates scholarly significance featuring the integration of theoretical aspects and applied studies. In view of the above, the thesis fully belongs to the subject area of the theory and practice of translation.

2. Particulars of the objectives and tasks of the Doctor of Science thesis

The objectives of the study are clearly stated only in the synopsis to the thesis. The major objective of the thesis is to explore the options of applying formal linguistic theories for the computational processing of the Bulgarian language and to Bulgarian machine translation in particular. The study is conducted by exploring the applicability of three approaches to Bulgarian

machine translation – DATR language for lexical knowledge representation, the universal networking language (UNL) for grammar representation and Sketch Engine (SE) for statistical processing of bilingual electronic corpora. The study tasks, that need to be solved in order to accomplish the research objectives, are not explicitly defined overlapping mostly with the study objectives.

3. Compliance of research methodology with the objectives of the thesis

The leading research methodology is contrastive analysis based on Tertium Comparationis and its application to the study of linguistic phenomena in two or more languages in order to outline similarities and dissimilarities between them. The relevance of machine translation to the larger area of translation studies is clearly stated. The features of texts suitable for machine translation are also defined. The traditions of machine translation research in Bulgarian linguistics and translation studies are highlighted. The major differences between professional human translation and machine translation are outlined in terms of richness of vocabulary and language structures, translation imagery, rendering culture specific vocabulary, etc. The two major approaches to translation technology – machine translation and computer-aided translation – are also dwelled on.

The study presents the stages of functioning of machine translation software in relation to encoding and decoding information in automatic translation from the source language into the target language. When retrieving the meaning of the source language, there is the intermediary stage of symbolic interlingual representation prior to rendering meaningful translation counterpart by using the structures of the target language. Machine translation employs mainly formal approaches of language processing. The three major approaches to machine translation are dwelled on in the thesis – rule-based machine translation, statistical machine translation and neural machine translation. Despite the fact that with the advent of neural networks the scope of texts for automatic translation has widened, the subject areas that are currently attainable only by professional human translators are clearly outlined in the study.

The thesis features a contrastive analysis of machine translation applications to processing of the semantic and grammatical peculiarities of the Bulgarian language. The selected research methodology complies with the objectives of the study as it includes a detailed presentation of the three approaches to Bulgarian machine translation – DATR, the universal networking language (UNL) and Sketch Engine (SE) in theoretical aspect and in view of their practical application. The

major approaches in contrastive analysis and machine translation are aptly integrated in the analysis and are applied to the study of linguistic phenomena in various languages such as the inflectional morphology in Bulgarian, Russian and Polish, the pronominal system in Bulgarian, English and Russian, etc. The thesis aims at presenting the applicability of language processing methods to contrastive analysis research, which is elucidated by applying a machine translation model between language pairs such as Bulgarian and Slovak, Bulgarian and Serbian, Bulgarian and English, etc. The practical aspects of the study are illustrated by analysing machine translation application to rendering certain linguistic units such as ecological terms and time expressions in Bulgarian and Slovak, etc. The presentation of computational approaches to terminology is significant for the study. Due to their monosemy, terms generally increase the success rate of machine translation output. As a whole, the thesis aptly demonstrates the integration of contrastive analysis methodology with machine translation approaches for the purposes of the study. The thesis synopsis demonstrates a comprehensive rendering of the issues related to the research topic following strictly the contents of the thesis with logical structure and smooth transition between various sections of the text.

4. Scholarly and applied practical contributions of the thesis

The thesis has a number of theoretical and applied practical contributions with reference to Bulgarian language processing and machine translation technology. The following major contributions can be highlighted: the application of current formal linguistic theories to a number of Bulgarian machine translation approaches; providing a pattern of metrics to evaluate Bulgarian machine translation output; elaborating on the recent advances in Bulgarian machine translation; establishing the feasibility of machine translation as an aid to professional human translators. By focusing on the Bulgarian language the thesis lays emphasis on languages with lesser number of speakers, thus highlighting the significance of machine translation technology without prioritising the number of speakers as a prerequisite for the development of translation technology. The thesis is written in English, which is also considered a contribution as the thesis might be brought to the attention of the global academic research community working in the area of translation technology and might popularise Bulgarian language machine translation systems.

5. Assessment of the publications related to the thesis

Dr Velislava Stoykova has authored a large number of scholarly publications mainly in the area of language technology and machine translation, published in specialised editions in Bulgaria

and abroad. The papers are written both in English and Bulgarian. For the last ten years, Dr Stoykova has over forty publications, some of which are referenced in Scopus. Twelve publications are presented as directly related to the topic of the thesis, some of which are also referenced in Scopus and Web of Science.

6. Citations

Dr Stoykova has over forty citations. A considerable portion of them are abroad, which is indicative of the international recognition of her scholarly work.

7. Comments and recommendations

The Bulgarian machine translation approaches presented in the thesis might be tested on specialised texts containing terminology such as economy, finance, healthcare, etc. Machine translation research might be applied extensively to the subject area of terminology. Some minor errors in the text should be avoided such as *the first* (p. 8) → *the former*, *later* (p. 8) → *the latter*, *underlay* (pp. 9, 10, 17) → *underlie*, *can regarded*, *can derived* (p. 9) → *can be regarded*, *can be derived*, *can be occurred* (p. 13) → *can occur*, *sentense* (p. 25) → *sentence*, *is represent* (p. 38) → *are represented*.

8. Conclusion and evaluation of the thesis

Dr Velislava Stoykova's thesis, scholarly publications and citations meet the scientific metric requirements of the Development of Academic Staff in the Republic of Bulgaria Act and its accompanying regulations. Considering that technological resources for Bulgarian have not been fully developed yet, the thesis has a marked contribution to the development of Bulgarian natural language processing and machine translation technology. I give my positive evaluation of the thesis and I would recommend the members of the esteemed Scientific Jury to vote positively for awarding the scientific degree Doctor of Science to Dr Velislava Stoykova in the field of higher education 2. Humanities, professional field 2.1. Philology. The Theory and Practice of Translation.

10.09.2022

Veliko Tarnovo

Prepared by:



/Assoc. Prof. Boryana Bratanova, PhD/