

REVIEW

By Prof. Irena Vassileva, PhD, NBU, scientific field 2.1. Philology

for the acquirement of the scientific degree "Dr.Sc." in professional field 2.1. Philology, candidate Velislava Raykova Stoykova, PhD

Topic of the thesis: „**Machine Translation Approaches to the Bulgarian Language**“

1. Scientific and applied significance of the research question.

The researched question is significant in the sense that the topic under consideration is very hardly represented both in theoretical and practical aspects regarding the problems of machine translation from and into Bulgarian. This applies even more to the highly developed inflectional morphology of Bulgarian and the other Slavic languages, which creates serious additional stumbling blocks for the design of machine translation software. Therefore, the results of the study are directly applicable to the theory and practice of machine translation.

2. Accurately formulated aims and objectives of the dissertation.

The objectives and tasks are clearly and precisely formulated both in terms of content and scope of the analysis, its results and their application.

3. Degree of knowledge of the state of the problem and relevance of the used literature.

The candidate is obviously well acquainted with the academic literature on the problem, since in the process of analysis, as well as in the theoretical overview, she relies on a sufficiently large number of authors. English-language sources predominate, and some Bulgarian ones are also included. At the same time, the approach to working with the sources is critical and demonstrates the author's ability to extract the essentials from the available publications in accordance with the specific needs of the research.

4. Correctness in referring to a representative number of authors.

The citations in the text are correct. The number of sources according to the Reference list is not large, but as it is clear from the theoretical review, the topic - the object of study - is quite narrow and underdeveloped.

5. Presence of a substantiated and developed theoretical model of the research.

The research is theoretically based on the basic principles of machine translation and the various theoretical and practical approaches to its implementation, combined with the application of various formal linguistic theories for automatic processing and analysis of the Bulgarian language. The choice of this model is convincingly justified by the lack of publications dealing with this issue.

6. Correspondence of the chosen methodology and research methods to the set aims and objectives of the dissertation.

For the purposes of the study, the method of contrastive analysis is used in a broader understanding of the concept, going beyond its use in purely linguistic studies. More specifically, three approaches to machine translation in Bulgarian are compared in terms of the presentation of linguistic data and their formalization, their application to machine translation, and the evaluation of the quality of the translation as a final product. Thus, the chosen methodology corresponds to the objectives of the study.

7. Presence of author's contribution to the collection and analysis of empirical data.

The collection of empirical data is from electronic databases and their processing is entirely carried out by the author.

Description of the contributions:

1. Brief description of the nature and assessment of the reliability of the material on which the contributions of the dissertation are based.

The materials which serve as the basis of the analysis in the study is carried out are extracted from a sufficient number of digital databases, whose reliability is beyond doubt. Different databases are selected depending on the specific tasks set in the individual parts of the research.

2. Description of the applicant's contributions and their classification.

The dissertation definitely demonstrates that the candidate has in-depth theoretical knowledge in the field of machine translation, formal grammar and the affordances of different machine translation languages with a focus on the Bulgarian language, as well as in the field of translation quality assessment.

The main contribution points in the dissertation are the delineation of the differences between the results of the machine translation when using three program applications for the Bulgarian language, accompanied by many examples, as well as the detailed description of their qualitative characteristics. Included in the analysis are two applications based on grammatical and semantic analysis and one application that primarily enables semantic analysis by creating concordances and semantic fields. For the purpose of the analysis, examples from Bulgarian, English, Russian, Slovak and Serbian are used, the Slavic languages being in the focus in view of their complex morphological characteristics.

It is here that I see the other contribution of the study, namely the methods for the formal presentation of the complex morphology of the Slavic languages and in particular of the Bulgarian language. Setting the source data in this sense is considered decisive for the quality of the translation.

The possibility of assessing its quality also contributes to the practice of translation - a problem facing both practicing translators and their trainers.

And last but not least, the author rightly, and based on evidence, establishes that the use of only one machine approach from the existing ones is not enough to achieve a sufficiently adequate translation, and that machine translation still has its limits and is difficult to implement to work with, for example, fiction texts due to the lack of possibility to include cultural context parameters.

3. Assessment of the degree of personal participation of the author in the contributions.

Although she relies on her previous (or co-authored) research on the problem, the candidate uses them skillfully and critically to support the new analyses and conclusions. Getting into even the smallest details of the semantic and syntactic features of the studied language elements and their translation equivalents shows that the author has put in a lot of efforts and worked independently.

4. Assessment of the correspondence between the extended abstract and the main points and contributions of the dissertation.

The extended abstract reflects clearly and accurately the content of the dissertation and its contributions.

Impacts of the dissertation on the external environment.

1. Evaluation of the publications on the dissertation: number, nature of the journals in which they are published.

Twelve publications, which concern individual aspects of the dissertation, published after 2011 are indicated. Four of them are co-authored. All publications are in English and are in international editions. This shows interest in the topic developed by the candidate in the last ten years. In addition, all articles are cited in the dissertation which clearly summarizes and expands the results of the author's research.

2. Citations by other authors, reviews in the scientific press, etc.

Eight citations are presented, seven of which are in publications referenced and indexed in recognized databases of scientific information or in monographs and collective volumes, and one - in a peer-reviewed monograph. Here I have to note that all the citations are by foreign authors, which, on the one hand, demonstrates the interest in the topic outside of Bulgaria and, unfortunately, the lack thereof in the country.

Personal qualities of the author (if the reviewer knows them).

I have no personal impressions.

Opinions, recommendations, and notes.

The dissertation focuses on machine translation tools of the "rules-based machine translation" generation, where rules are required, and "statistical machine translation", based on data extracted from bilingual translation corpora. At present, however, neural machine translation technology is considered much more effective, and is used by practically all the world's leading machine translation companies, such as: Google Translate, DeepL Translator, Bing Microsoft Translator, SYSTRAN Translate, Amazon Translate, to mention only the top five in terms of influence and distribution. Therefore, I believe that the dissertation should not only mention these tools, but also explain why (if this is the case) they are not more adequate for application to the Bulgarian language.

Overall, the paper references exclusively both publications and machine translation approaches that support the proposed analysis and results. Other approaches used for the Bulgarian language are not mentioned, either. Situating the study in a wider context that demonstrates a critical view of alternative theories and practices would make it sound more convincing and credible. In this sense, I recommend that the author expand the overview part by supplementing it with a critical analysis by pointing out the advantages of her approach over others.

Conclusion with a clearly formulated positive or negative assessment of the dissertation.

My overall assessment of Dr. Velislava Raykova Stoykova's dissertation, extended abstract and publications is positive. The candidate is a proven, internationally recognized scholar in her field. The dissertation contains both theoretical and applied contributions in the practice of machine translation and, in particular, in solving the problems arising in the formalization of Bulgarian grammar for the purpose of its automatic processing. The structure and content of the submitted dissertation and abstract fully meet the requirements of the LDASRB and the regulations for its implementation at NBU. This gives me a reason to propose that the scientific degree "Doctor of Sciences" be awarded to Dr. Velislava Raykova Stoikova in professional field 2.1. "Philology, Scientific specialty *Theory and practice of translation*".

Date: 10.08.2022

Signature: Prof. Irena Vassileva, PhD