ASSESSMENT OF A THESIS FOR THE AWARD OF THE ACADEMIC DEGREE "DOCTOR"

By Associate Professor Elena Borisova Tarasheva, PhD in Mathematical Linguistics, New Bulgarian University, habilitated in the field 2.1. Philosophy, Media Discourse
For the applicant for the degree of Doctor in the professional field 2.1. Philology: Theory and Practice of Translation Velislava Raykova Stoykova, PhD.

1.   Significance of the research problem in scientific and applied terms.

The thesis attempts to evaluate the output of three formal linguistic theories (FLTs) for automatic processing and analysis of Bulgarian in translation from/to several Slavic languages and English. The marking of several different morphological categories is reviewed with a focus on machine translation.
Natural language processing for machine translation purposes is a popular field in modern linguistics. This is justified by the increased use of machine-assisted translation. At the same time, the study of natural languages is rapidly growing in view of the social demands for fast, accurate and adequate translation not only of technical translations but also of socio-political and texts from other genres. Looking at directions in linguistics other than structural linguistics - popular in the last century in many Slavic countries and the rising interest in functional linguistics, generative grammar, statistical methods, and formal grammars is much needed and timely in Bulgarian linguistics.
Research at the level of syntax in relation to machine language processing for the Bulgarian language marks a serious growth, in contrast to investigating the morphological structure of the lexicon. Engaging specifically with categories such as number, gender and definiteness is a necessary addition to existing projects.
Especially important is the practical value of the research work. Machine translation has long been a widely used tool in the translation community rather than a "crutch" for those unfamiliar with the language. The intensive use of machine translation in documents of the European Community, the United Nations and other international organisations, as well as for technical, medical and other genres implies a serious and responsible attitude towards this translation option, backed by research. The thesis presented by the candidate offers precisely such an effort and thus deserves high praise for its practicality and for its contemporary attitude towards linguistics.
I find the application of semantic networks particularly relevant to machine translation. In this work it is applied to DATR on Slavic languages, which are typically characterized by extended lemmas of linguistic expressions in relation to definiteness, gender, and number. While the review of types of machine translation explicitly names the statistical and rule-based approaches, the work implicitly includes the newest one, neural nets, which does exploit the morphological nature of lexical units.

2.   Justification of the aims and objectives of the dissertation.
The doctoral candidate sets herself the task of evaluating the formal representation of the Bulgarian language with the help of three linguistic theories using contrastive analysis. Assessment of the adequate representation of the morphological forms of linguistic units by the formal linguistic theories chosen in the paper is necessary for machine translation purposes. Such an assessment employs linguistic knowledge from structural linguistics and takes it to a new level, as well as conditioning modern applications such as machine translation.
I believe that the candidate has set herself ambitious but realistic goals, tested material on applications that find widespread use in reality, such as machine translation, which underpins all types of translation nowadays. The appropriate research approach to data analysis has been adopted and the likelihood of the analysis producing good results is sufficiently plausible.

3.   Conformity between the chosen methodology and research methodology and the set aim and objectives of the dissertation.
The chosen model enables the candidate to compare how different morphological categories are represented within

Образец за структура на становище на научен труд за придобиване на образователната и научна степен „доктор" и научната степен „доктор на науките" в НБУ

Приложение 5 към Наредба за развитието на академичния състав на НБУ

the three formal theories, thereby allowing practitioners of the respective levels of formalisation to select the one that best suits their purposes. The comparison is not one-dimensional as the formal theories work with different categories. Different languages are also used for comparison: Russian, Slovak, Polish, English etc. With this, the analysis demonstrates a level of high abstraction and purposefulness befitting the highest academic title.

In most cases, the results of the study justify the preliminary hypotheses, which demonstrates a mastery of the methodology at a level appropriate to the requirements of the advanced degree being defended.

4.   Theoretical and applied contributions of the dissertation (description and evaluation), including the presence of original contributions to science.

As her main contribution, the candidate reports her analysis of the approaches to computer-assisted translation in Bulgarian. The distinction of types of translation equivalence is also highly appreciated. As a particularly valuable issue she deems the automated evaluation of machine translation results. Not to be forgotten is the training corpus of mathematical texts for translation.

The serious approach to machine translation and the software products for its creation, in my opinion, represent the primary contribution to the broadening of the horizons of linguistics and its orientation towards everyday practical problems in life.

The consideration of morphological categories in relation to translation has also not received sufficient attention within formal grammars at the present time. Thus, I find the candidate's orientation to morphology to be contributory and especially beneficial given the nature of Slavic languages and the wide application of inflectional categories.

5.   Assessment of the publications on the dissertation: number, nature of the publications in which they have been published.

As required by the degree of Doctor of Science, the candidate has a significant number of publications, most of which feature in refereed journals. In the database Research Gate the interest index of the candidate's profile is 28.8 - a high value for a humanities scholar, 73 citations of her work in academic publications are listed. The H-index of publications in Scopus is also 4, with citations peaking twice for the years from 2018 to the defense of this thesis. Twenty-three articles are recorded in the prestigious scientific database Scopus, which few colleagues from Bulgaria have achieved.

6.   Opinions, recommendations and comments.

Calling Sketch Engine a "search engine" does not fit the way the product is presented by its creators. They call it a linguistic software product, an analytical tool, etc., which also reflects its specificity: it does not just search for occurrences of selected forms, but analyses collocates, frequencies, mutual information, genre characteristics and other features, which is why it is better to use a more sophisticated term than "search engine".

The presentation of machine translation into statistical and rule-driven is simplistic and does not reflect current trends. Even the most traditional translation tool Google has adopted neural networks to replace the statistical model. Therefore, the neural model deserves more substantial discussion in the thesis, especially given the semantic focus of the morphological categories under investigation.

Ossenova and Simov's Formal Grammar of the Bulgarian Language is briefly mentioned in the bibliography, as are authors such as Paskaleva, Lyudskanov, Legurska, Ivanchev, etc. Their work is directly relevant to formal grammars, translation, and digital approaches to linguistics. However, I find it a weakness not to include Koeva in the review, given the serious attempts she has made in terms of translation products, language formalization, and especially the morphological structure of the lexicon.

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

However, the weaknesses noted here are correctable and technical. They do not detract from the value of the work. I therefore recommend that the degree of DOCTOR OF SCIENCE be awarded to Velislava Stoykova.

Образец за структура на становище на научен труд за придобиване на образователната и научна степен „доктор" и научната степен „доктор на науките" в НБУ

Приложение 5 към Наредба за развитието на академичния състав на НБУ

Date…………… September, 15th 2022   Signed……………