

Review

from

Prof. Dr. Stefka Fidanova

Institute of Information and Communication Technologies

Bulgarian Academy of Sciences

expert in the field of knowledge 4. 6. Informatics and Computer
Sciences

for

acquisition of the scientific degree of “Doctor of Science” in the
field of knowledge 2.1 Philology by the candidate Dr. Velislava
Raykova Stoykova.

I have been approved by order No. 3 - RK - 268/13.07.2022 of the Rector of the New Bulgarian University as a member of a scientific jury for the procedure of acquisition of the scientific degree of “Doctor of Science” in the field of knowledge 2.1. Philology by Dr. Velislava Raykova Stoykova with a thesis “Machine Translation Approaches to the Bulgarian Language“ (Подходи за машинен превод на български език).

As a member of the Scientific jury, I have received:

1. The dissertation for the acquisition of a scientific degree “Doctor of Science” in the field of knowledge 2.1. Philology;
2. The resume of the dissertation;
3. The professional CV;

4. The list of author's publications associated with the dissertation.

The requirements of the Law on Development of Academic Staff in the Republic of Bulgaria (LDASRB), the Rules on its Implementation (RILDASRB), the corresponding Rules of Procedure, and the Ordinance on the development of academic staff of the NBU shall be taken into account when assessing the dissertation work.

- Under Art. 6 (3) of the LDASRB “the dissertation must contain scientific or applied results which represent an original contribution to science. The dissertation work must show that the applicant has in-depth theoretical knowledge of the relevant field and capability for independent research work”.
- According to Art. 27 (2) of the RILDASRB the dissertation shall be provided in a form and volume corresponding to the specific requirements of the primary academic unit. The dissertation work must contain a title page; content; introduction; exhibition; conclusion – summary of results obtained; bibliography.

The proposed dissertation consists of an introduction, 3 chapters, a conclusion, bibliography, appendices, and reference apparatus (lists of abbreviations, index of main terms, etc.). The list of the author's scientific publications is presented, a list of the related quoted citations, and the projects in which the candidate has participated are presented as well.

1. Relevance of the problem and appropriateness of the objectives and targets set

We live in times when computers influence all spheres of our lives. The turbulent development of many areas in recent decades has been due to the penetration of computer support. This leads to the need for rapid and high-quality translation. The development of language technologies supports the development of machine tools for machine translation. The development of such technologies begins, a few decades ago, with the development of English. Unfortunately, the research in this field, linked to the Bulgarian language, is still in the beginning. This makes the proposed dissertation work particularly up-to-date.

The presented to me for the evaluation work is mainly aimed at analyzing the existing approaches in the field of language technologies and their capabilities and effectiveness for their use in the Bulgarian language. Thus, the main objective of the entire dissertation work is to explore and analyze the effects of the use of formal language technologies for computer analysis of the Bulgarian language and mainly their ability to be used for machine translation.

In order to achieve the objective set, the following tasks are defined:

- Examination of a variety of approaches for machine translation for the Bulgarian language using a contrastive analysis.
- Comparison of computer programs for Bulgarian and the results of their work, taking into account the presentation of language fragments; the time and resources they use; the quality of the received translation, and the opportunities for improvement and development.

2. Knowledge of the status of the studied problems in the related field by the applicant

There is no doubt that the applicant has very good knowledge and experience in the field of language technologies and the related theoretical and applied problems. Almost 30% of the presented bibliographic sources quoted in the dissertation are over the last 10 years, making this work very up-to-date. On the other hand, it also includes references to older sources which are key to studied issues. The total number of cited sources is 107, with 89 in English, 16 in Bulgarian, and 2 in Russian.

The applicant's good knowledge and experience in studying the problem are demonstrated and very well illustrated in Chapter 1 of the dissertation. There are detailed analysis and the general assumptions for the machine translation field.

3. Study methodology

The methodology for conducting the research chosen by the applicant results from the objective set and corresponds to the tasks arising therefrom. It is perfectly adequate for the subject studied in the work. The author uses the DATR language for lexical knowledge representation, the Universal Networking Language, and the Sketch Engine in the study. These approaches have been developed mainly for English. The author considers their capabilities and the necessary extrapolations and further development so that they can be successfully implemented for processing the Bulgarian language.

4. Characteristics and evaluation of the contributions of dissertation work

The dissertation starts with an introduction to the subject area of the study. There is a motivation to select the problem and the subject of the study. A brief overview of the methods used in the dissertation was made. The objectives and tasks of the dissertation are specified. Tables and figures illustrating the tasks in question and the results achieved are attached.

Chapter 2 presents the formal DATR language and its potential for the application for machine translation in Bulgarian and necessary further development in order to be effective.

Chapter 3 is dedicated to the Universal Networking Language and its use for machine translation. It is used for the presentation of various language phenomena in Bulgarian. A comparison was made between English and Russian, showing the machine translation mechanism between them. Examples of machine translation between Bulgarian and Slovak languages using Universal Network Language are also presented.

Chapter 4 addresses a search method in electronic bilingual databases using a statistical approach. Statistical approaches provide good results when searching in large databases, in particular parallel and comparable bilingual electronic text corpora. They are fast and at the same time, give good quality translation results. The approach used is supported by examples of machine translation in Bulgarian, Slovak, and Serbian languages.

The dissertation has noted the following contributions:

1. The detailed analysis of the Bulgarian language translation approaches based on the use of the formal language theories of the DATR language and the Universal Networking Language, their respective programming applications as well as the detailed analysis based on the statistical approaches of the Sketch Engine are given.
2. The correspondence between semantic equivalent, logical equivalent, and translation equivalent is accepted as the basis of the machine translation processing of the Bulgarian language.
3. In the analysis and evaluation of the machine translation processing of the Bulgarian language, both the starting and the received linguistic data are used, which are supported by convincing examples.
4. The approach to measure the accuracy of the translation of the Bulgarian language is given, as well as the approach to improve the quality of the received translation

Guidelines for possible future development of the topic are given.

5. Significance of the study for future scientific and practical development

The work carried out by the applicant is sufficient in terms of volume and depth of the study. The approaches shown are of practical importance. In this sense, I find the work important both in scientific and practical terms.

6. Estimation of dissertation publications

In connection with the dissertation, the applicant submitted 12 publications. 7 of them are in the editions that are indexed in the world global system of scientific referring and indexing, with 4 being authored by the applicant alone. 5 are in peer-reviewed non-referenced publications, with 4 of them being authored by the applicant alone. The total number of points by a group of indicators Γ is 200 (of 100 required). According to indicators of group Δ , Stoykova presented 7 citations, in referenced editions of its publications, whereby the total number of points was 105 (of 100 required). All publications and citations are after the applicant has received the educational and scientific degree of "Doctor". This is proof that they have not been used in previous procedures.

In my opinion, the presented to me dissertation meets the requirements of the LDASRB, as well as the related rules for its implementation by presenting an authoritative study on a significant scientific problem – machine translation, which has substantial scientific and practical results. The author of the dissertation corresponds to and even exceeds the required parameters in terms of publications and citations presented to the dissertation.

7. Applicant's personal participation

8 of the publications presented in the dissertation are self-authored, which clearly shows the personal participation of the candidate.

8. Resume of the dissertation

As a whole, the resume of the dissertation correctly reflects the content of the dissertation itself.

9. Critical notes

The tasks of the dissertation are not strictly formulated in its resume. Although the resume of the dissertation correctly reflects its content and its contributions, it does not follow exactly the content of the dissertation and its contributions and does not completely follow the structure of the dissertation.

10. Personal impressions

I know Velislava Stoykova for a long time. She is a dedicated scientist who successfully combines her knowledge in the field of Bulgarian language and new computer technologies. She is one of the first in Bulgaria who introduced the use of computer technology for the analysis and processing of texts. Velislava Stoykova was successful, as evidenced by its publications, in participating in philology conferences and conferences in the field of applied mathematics and informatics. This shows how fruitful it can be to combine these scientific areas.

11. Conclusion

As a consequence of the above, it can be found that all the requirements of the Law on Development of Academic Staff in the Republic of Bulgaria (LDASRB), its Implementing Regulations (RILDASRB), the corresponding Rules of Procedure, and the Ordinance on the development of academic staff of the NBU are met. I can say that the level of this dissertation satisfies and significantly exceeds the minimum requirements.

The misstatements and observations made by me do not reduce the significance of the results obtained and the scientific value of the work provided to me.

All of this gives me the basis for a positive assessment, and I propose to the Honorable Scientific Jury to award the scientific degree “Doctor of Science” in the field of knowledge 2.1. Philology to Velislava Raykova Stoykova.

25.07.2022,

Sofia

Prof. Dr. Stefka Fidanova