

О Т З И В

От проф. д-р Ирена Георгиева Василева, НБУ, научно направление 2.1. Филология за придобиване на научната степен „доктор на науките“ по професионално направление 2.1. Филология, с кандидат д-р Велислава Райкова Стойкова

Тема на дисертацията: **„Machine Translation Approaches to the Bulgarian Language (Подходи за машинен превод на български език)“**

1. Значимост на изследвания проблем в научно и научно-приложно отношение.
Изследваният проблем е значителен в смисъл, че разглежданата тематика е много слабо представена както в теоретичен, така и в практически аспект по отношение на проблемите на машинния превод от и на български език. Това се отнася с още по-голяма сила за силно развитата словоизменителна морфология на българския и другите славянски езици, която създава сериозни допълнителни спънки при създаването на софтуер за машинен превод. Ето защо резултатите от разработката са директно приложими в теорията и практиката на машинния превод.
2. Точно формулирани цели и задачи на дисертацията.
Целите и задачите са ясно и точно формулирани както като съдържание, така и като обхват на анализа, неговите резултати и тяхното приложение.
3. Степен на познаване на състоянието на проблема и съответствие на използваната литература.
Кандидатката очевидно е добре запозната с научната литература по проблема, като в процеса на анализ, както и в теоретичния обзор се опира на достатъчно голям брой автори. Преобладават англоезични източници, като са включени и български. Същевременно, подходът при работата с източниците е критичен и показва уменията на авторката да извлича същественото от наличните публикации в съответствие с конкретните нужди на изследването.
4. Коректност при цитирането на представителен брой автори.
Цитиранията в текста са коректни. Броят на източниците според Библиографията не е голям, но както става ясно от теоретичния обзор, темата – обект на изследване е относително тясна и малко разработвана.
5. Наличие на обоснован и разработен теоретичен модел на изследването.
Изследването се основава теоретично на основните принципи на машинния превод и различните теоретични и практически подходи към неговата реализация, съчетани с приложението на различни формални езикови теории за автоматична обработка и анализ на български език. Изборът на този модел е обоснован убедително с липсата на публикации, третиращи тази проблематика.
6. Съответствие на избраната методология и методика на изследване с поставената цел и задачи на дисертационния труд.

За целите на изследването се използва метода на контрастивния анализ в едно по-широко разбиране на понятието, надхвърлящо употребата му в чисто езиковедски изследвания. По-конкретно се сравняват три подхода за машинен превод на български език по отношение на представянето на езиковите данни и тяхното формализиране, тяхното приложение за машинен превод и оценяването на качеството на превода като краен продукт. Така избраната методология съответства на целите на разработката.

7. Наличие на собствен принос при събирането и анализирането на емпиричните данни. Събирането на емпиричните данни са от електронни бази данни и тяхната обработка е изцяло извършена от авторката.

Описание на приносите:

1. Кратка характеристика на естеството и оценка на достоверността на материала, върху който се градят приносите на дисертационния труд.
Материалите, въз основа на които се извършва анализа в разработката, са извлечени от достатъчен брой дигитални бази данни, в чиято достоверност няма съмнение. Подбрани са различни бази данни в зависимост от конкретните задачи, поставени в отделите части на работата.
2. Описание на приносите на кандидата и класифицирането им.
Дисертационният труд определено показва, че кандидатката притежава задълбочени теоретични знания в областта на машинния превод, формалната граматика и възможностите на различни машинни езици за превод с фокус върху българския език, както и в областта на оценката на качеството на превода.
Основни приносни моменти в дисертацията са очертаването на разликите между резултатите от машинния превод при използването на три програмни приложения за българския език, придружени с много примери, както и детайлното описание на качествените им характеристики. В анализа са включени две приложения, базирани на граматичен и семантичен анализ и едно приложение, което позволява предимно семантичен анализ чрез създаване на конкорданси и семантични полета. За целта на анализа се използват примери от български, английски, руски, словашки и сръбски, като славянските езици са основни с оглед на техните сложни морфологични характеристики.
Именно тук виждам другия принос в разработката, а именно методите за формално представяне на сложната морфология на славянските езици и в частност на българския език. Задаването на изходните данни в този смисъл е решаващо за качеството на превода.
Към практиката на превода допринася и предложената в работата възможност за оценка на качеството му – проблем, който стои както пред практикуващите преводачи, така и пред техните обучители.
И не на последно място, авторката с право и въз основа на доказателства установява, че използването само на един машинен подход от съществуващите засега не е достатъчно за постигане на достатъчно адекватен превод, както и че машинният превод все още има своите граници и е трудно приложим към работа

с, например, художествени текстове поради липсата на възможност за въвеждане на параметри на културен контекст.

3. Оценка на степента на лично участие на дисертанта в приносите.

Въпреки че се опира на предишни свои (или в съавторство) изследвания на проблема, дисертантката ги използва умело и критично, за подкрепа новите анализи и изводи. Навлизането дори и в най-дребните детайли на семантичните и синтактични характеристики на изследваните езикови елементи и техните преводни еквиваленти показват, че авторката е положила много труд и е работила самостоятелно.

4. Оценка на съответствието на автореферата с основните положения и приносите на дисертационния труд.

Авторефератът представя ясно и точно съдържанието на дисертацията и нейните приноси.

Въздействия на дисертационния труд върху външната среда.

1. Преценка на публикациите по дисертационния труд: брой, характер на изданията, в които са публикувани.

Посочени са дванадесет публикации, които засягат отделни аспекти на дисертационния труд, публикувани след 2011 г. От тях четири са в съавторство. Всички публикации са на английски език и са в международни издания. Това показва интерес към тематиката, разработвана от докторантката през последните десетина години. Освен това всички статии са цитирани в дисертацията, която очевидно обобщава и разширява резултатите от научните търсения на авторката.

2. Използване и цитиране от други автори, отзиви в научния печат и др.

Представени са осем цитирания, седем от които в научни издания, реферирани и индексирани в световноизвестни бази данни с научна информация или в монографии и колективни томове и едно – в монография с научно рецензиране. Тук трябва да отбележа, че всички цитирания са от чужди автори, което от една страна говори за интереса към разработената тематика извън България и, за съжаление, липсата на такъв в страната.

Лични качества на автора (ако рецензентът го познава).

Нямам лични впечатления.

Мнения, препоръки и бележки.

Дисертацията се фокусира върху инструменти за машинен превод от поколението „rules-based machine translation“, където се изисква задаването на правила, и „statistical machine translation“, основаващи се на извлечени данни от двуезични преводни корпуси. Към настоящия момент, обаче, за много по-ефективна се смята технологията на „neural machine translation“, която се използва от практически всички водещи в света компании,

предлагачи машинен превод, като: Google Translate, DeepL Translator, Bing Microsoft Translator, SYSTRAN Translate, Amazon Translate, ако споменем само първите пет по влияние и разпространение. Ето защо мнението ми е, че в дисертацията би трябвало тези инструменти не само да се споменат, но и да се обясни защо (ако това е така) те не са адекватни за приложение за българския език.

Като цяло в работата се реферира изключително както към публикации, така и към подходи за машинен превод, които са в подкрепа на предложени анализ и резултати. Не се споменават също и други подходи, използвани за българския език. Ситуирането на разработката в един по-широк контекст, в който се демонстрира критичен поглед върху алтернативни теории и практики, би я направила да звучи по-убедително и достоверно.

В този смисъл препоръчвам на дисертантката да разшири обзорната част, като я допълни с критичен анализ и посочи предимствата на избория от нея подход пред други такива.

Заключение с ясно формулирана положителна или отрицателна оценка на дисертационния труд.

Общата ми оценка за дисертационния труд, автореферата и научните публикации на д-р Велислава Райкова Стойкова е положителна. Докторантката е доказан, международно признат учен в своята област. Дисертационният труд съдържа както теоретични, така и приложни приноси в практиката на машинния превод и в частност при решаване на проблемите, възникващи при формализирането на българската граматика с цел неговата автоматична обработка. Структурата и съдържанието на представените дисертационен труд и автореферат напълно отговарят на изискванията на ЗРАСРБ и правилника за прилагането му в НБУ. Това ми дава основание да предложа да бъде присъдена научната степен „Доктор на науките” на д-р Велислава Райкова Стойкова в професионално направление 2.1. “Филология, Научна специалност *Теория и практика на превода*”.

Дата 10.08.2022

Подпис: проф. д-р Ирена Василева