

РЕЦЕНЗИЯ

на дисертационен труд за придобиване на образователната и научната степен *доктор*, област 2. Хуманитарни науки, професионално направление 2.1. Филология, докторска програма „Общо и сравнително езикознание“.

Автор на дисертационния труд: Тодор Димитров Лазаров, докторант към Секцията по компютърна лингвистика на Института за български език „Проф. Л. Андрейчин“ – БАН.

Тема на дисертационния труд: „Българско-английски граматични паралели с оглед на машинния превод“.

Рецензент: проф. д-р Красимира Славчева Алексова, Катедра по български език, Софийски университет „Св. Климент Охридски“, член на научното жури със заповед РД-09-10 от 07.02.2020 г. на Директора на Института за български език проф. Светла Коева.

1. Актуалност на разработвания в дисертационния труд проблем в научно и научно-приложно отношение.

Прегледът на документите по процедурата показва, че са изпълнени Минималните национални изисквания за придобиване на научната и образователна степен „доктор“, както и изискванията на Правилника за докторанти в БАН.

Темата на дисертационния труд безспорно е дисертабилна и актуална, което се подкрепя от анализа на постигнатите резултати в изпълнение на поставените цели и задачи. Българско-английските паралели на времената и като система, и като употреби сами по себе си представляват самостоятелна дисертабилна тема, а в рецензирания труд тези теоретични наблюдения са последователно и целенасочено подчинени на една съвременна за компютърната лингвистика задача – подобряване на машинния превод. Резултатите от разработките на дисертанта имат неоспорима практическа приложимост, тъй като могат пряко да служат за подобряване на машинния превод от български на английски език.

2. Познава ли дисертантът състоянието на проблема и научната литература.

Трудът показва много добро познаване на теоретичните възгледи (както на български учени, така и на утвърдени в лингвистиката чужди изследователи), засягащи граматичните глаголни категории, които са от значимост за разработката. При това задълбочените познания на Т. Лазаров се разпростират в две области, които са взаимопроникващи се в дисертационния труд: теоретична граматика на българския и на английския език с акцент върху граматичните категории на глагола, от една страна, и от друга страна – машинния превод, който е част от широката област на компютърната

лингвистика. Прави впечатление, че концепциите на отделните учени са представени не самоцелно, а чрез сравнителен анализ, като освен това в работата се извеждат значими тенденции в развитието на машинния превод.

Библиографията обхваща над 150 цитирани единици, които са достатъчно основание да се представи една ясна картина на научните изследвания, свързани с глаголната система в двата съпоставяни езика, а и с машинния превод.

3. Избраната методика на изследване може ли да даде отговор на поставените в дисертационния труд цели и задачи.

В изпълнението на целите и задачите на дисертационния труд Т. Лазаров е приложил напълно адекватна методика. На базата на възприетите теоретични позиции за семантиката и състава на темпоралните системи на българския и английския език се прави формално-семантично описание на темпоралните значения в двата езика. Това става база за създаване на хипотези за преводните съответствия. Те се проверяват чрез емпирични данни от създадения въз основа на формулирани критерии Българско-английски корпус със съотнесени глаголни форми. Това означава, че корпусният подход се използва за верифициране на изследователските хипотези на докторанта. От съществена значимост са и статистическите анализи на темпоралните форми в двата езика и съпоставителният анализ по отношение на трансфера на информация за морфологичните категории, дефинирани от докторанта като време, вид на действието и таксис в българския, към английските морфологични категории време и вид на действието, а и за трансфера на информацията за категориите лице, число и род на българския глагол при превод на английски език.

4. Кратка аналитична характеристика на естеството и на достоверността на материала, върху който се градят приносите на дисертационния труд.

Тъй като е приложен корпусният анализ, може да се твърди, че достоверността на материала е гарантирана. Разбира се, анализите и изводите се базират на създадения Българско-английски корпус със съотнесени глаголни форми, който има определен (краен) обем, извън който могат да останат някои случаи на редки употреби, напр. на бъдеще предварително и бъдеще предварително време в миналото в българския език. Именно това се е опитал да избегне дисертантът, като е включил в Корпуса примери от граматиката на Р. Ницолова, издадена на английски език.

5. В какво се заключават научните или научно-приложните приноси на дисертационния труд и каква е по-конкретно значимостта на приносите за науката и практиката.

Дисертационният труд се отличава с логическа последователност на изложението, има ясна структура и наличие на връзки между отделните части. Обособените две макрочасти, посветени на лингвистичната теория и на машинния превод, са обвързани

целенасочено така, че се гарантира максималната реализация на основната цел на изследването – извличане на конкретни граматични паралели между темпоралните отношения, открити в паралелен българско-английски корпус и създаването на статистически модели за превод на глаголните форми, които да подобрят машинния превод. Петата глава напълно убеждава, че резултатите от дисертационния труд имат пряко практическо приложение – те позволят да се усъвършенства машинният превод. Напълно убедено мога да кажа, че теорията и практиката в изследването са максимално обвързани.

Според моята гледна точка успешността на реализация на поставените цели зависи до голяма степен от избора на теоретична позиция за организацията на категорията *време*. Това се осъзнава и от самия докторант, както личи от записаната основна цел на дисертацията на с. 3, и по-точно – на първата част от тази цел – „да се направи съпоставка на теоретично равнище между семантиката на темпоралните значения в двата езика, което ще подпомогне теоретичното обосноваване на възможните преводни еквиваленти ...“. Докторантът се опира на теоретичните възгледи на Иван Куцаров за граматичните категории на българския глагол, от една страна, а от друга – застъпва виждането за 9 времена в българския език, което влиза на пръв поглед в известно противоречие. Годор Лазаров го преодолява, като приема, че категорията време е хиперкатегория и представлява съвкупност от три морфологични категории в българския език според класификацията на глаголните граматични категории на Ив. Куцаров: време, вид на действието и таксис. Макар че този възглед (за трите отделни категории по виждането на Ив. Куцаров) може да бъде оспорван, опирането на виждането, че граматическото време е хиперкатегория, дава общата основа, която разрешава на докторанта да търси паралели между две различно организирани от типологическа гледна точка темпорални системи – българската и английската.

Несъмнено изборът на теоретична лингвистична позиция, от която да се изхожда в едно съпоставително по своята същност изследване, не е лесна задача при наличието на разнородни възгледи не само в граматическите традиции на двата езика, но и в типологията, като непременно трябва да се отчетат като трудност и разликите в набора от граматични глаголни категории в двата езика и нееднаквостта в състава и организацията на глаголните категории. Докторантът е показал възможност да извлече положителното от нееднородните позиции с цел обосноваване на своята теоретична концепция и еднозначното дефиниране на използвания терминологичен апарат.

Положителна черта на избрания от докторанта подход е отчитането на взаимодействието между граматичните и лексикалните елементи за изразяване на темпоралност, защото без това не може да се търси оптимизиране на преводните съответствия. Все пак граматичните категории функционират в текста чрез своите употреби в конкретния контекст във взаимодействие с лексикалните елементи. Реализацията на българските времена в свързания текст е подчертано контекстуално зависимо, особено в сложното изречение, за разлика от съгласуването на времената в

английския език. Този факт предопределя разнородността на преводните съответствия на българските глаголни времена в английските текстове. В дисертационния труд се повдига този съществен въпрос, който пряко засяга реализацията на една от изследователските цели, свързани с преводните съответствия на темпоралните форми – това е правилото за съгласуване на времената (с. 61) в английския език.

Много важни за реализиране на целите на дисертационния труд са две хипотези, които авторът подлага на проверка чрез извличането на статистическа информация от паралелния българско-английски корпус, а това са: дали съществува някакво съответствие между вида на глагола в българския и продължителните глаголни форми в английския (така от употребите може да се върви към теоретичното обобщение дали може в крайна сметка да се говори за вид в широк смисъл, който в езиците се граматикализира по различен начин – чрез вида в славянските езици и чрез продължителни глаголни форми, като по принцип това са различни семантични явления от по-общо в семантичен план поле), и втората хипотеза – съществува ли съответствие между аорист и минало просто в английския и имперфект и минало продължително време при отчитане на видовете различия в българския език (т.е. по емпиричен път се търсят доказателства за граматикализирането по различен начин на сходни семантични признаци) (с. 81). Тези хипотези са оформени като работни въпроси в т. 6 на с. 84.

Докторантът изчерпателно е представил историческия развой на идеите и подходите в машинния превод (както традиционните статистически системи за машинен превод, така и машинния превод, базиран на невронните мрежи). Необходима част от дисертационния труд е и разглеждането на характеристиките на езиковите корпуси като ресурс за машинния превод, тъй като достатъчният емпиричен материал и прилагането на значими критерии към корпуса гарантират адекватността на изследването.

Смятам, че в дисертацията корпусът, с който работи докторантът – Българско-английски корпус със съотнесени глаголни форми, е достатъчно ясно представен, като са изяснени неговите параметри. Дискутирани са проблемите в отношенията между критериите балансираност и илюстративност на използвания корпус поради наличие на нискочестотни глаголни форми. Авторът сам подчертава, че въпреки представителността и илюстративността на езиковите данни в Корпуса остава проблемът за граматическото разнообразие. В отделна част са представени и анотационните конвенции (два анотационни слоя) при създаването на използвания корпус.

Статистическите данни за разпределението на българските глаголни времена в използвания корпус, абсолютните стойности и процентните съотношения в четири типа текстове (художествена литература, публицистични текстове, научни текстове и филмови субтитри) дават безспорни основания за формулиране на твърдения, които доказват изказвани в научната литература хипотези, базирани се на изследователската интуиция. Подобни твърдения се отнасят до честотата на употребите на времената според степента на тяхната семантична маркираност, заменянето на по-маркираните времена с техните немаркирани корелати, зависимостта на честотата на употребата на времената от типа на

текста и др. Съпоставката между разпределението на българските и на английските времена в двете части на корпуса (българска и английска) дава основание за подкрепа на вижданията, че честотата на употребите следва семантичната натовареност на членовете на една категория, а разпространението им е зависимо и от характерологичните стилкови особености на текста.

От гледна точка на научните ми интереси Таблица 15 с разпределението на преводните съответствия на формите на българското сегашно време в английските текстове е показателно, особено що се отнася до положението в художествените текстове и филмовите субтитри, защото причините не се състоят само в това, че има най-много транспозиционни употреби, както сочи авторът, но и във факта, че в българския език граматическите категории се отличават с повишена степен на контекстова зависимост – ако контекстът е изразил съответното значение, в случая – темпорален признак, то маркираното с този признак време може да бъде заменено с неговия немаркиран корелат. Това важи не само за 1, но за 2 и за 3 признака, което води до замяната дори на бъдеще предварително време от сегашно време в българския език. Таблица 15 е пряко доказателно за този факт. Също такива доказателства се съдържат и в Таблица 20, която показва, че на сегашни форми в българския език в 32.4% от случаите съответстват минали времена в английския и 5.92% бъдещи времена.

При минало свършено време, въпреки наличието на няколко преводни съответствия, вариатността е по-малка, което се дължи на неговата семантична натовареност с признака предходност спрямо акта на комуникация. Струва ми се, че случаите с единични преводни съответствия се дължат на семантични и/или видови особености на българския глагол. Данните от Таблица 19 с разпределението на преводните съответствия на българското минало несвършено време според моите виждания са пряко доказателство, че двете времена имат различни места в организацията на темпоралните системи на двата езика и не покриват една и съща семантика. Данните от преводните съответствия на българския перфект в английските текстове биха били достоверни, ако не са кодирани като перфект употребите на конклузивния аорист. Преобладаването на преводни съответствия с минало просто в английския език доказва различия не толкова в организацията на категорията време, колкото във функционирането на времената, вкл. в сложното изречение.

Изводът, че категорията таксис в български няма строго определени преводни еквиваленти в английския, макар почти 60% са преводни съответствия с прости времена в английския, има своята причина, която се корени според мене във факта, че в английския език функционирането на времената се управлява от съгласувателни правила, докато реализацията на индиректните времена (наречени зависим таксис) се подчинява на контекстуалната зависимост, а не на obligatorността.

Важна значимост за езиковедската теория и типологическите изследвания има доказаната чрез корпусния подход теза, че няма пряка връзка на формално ниво, а това е значимо и за носената семантика, между категорията вид на глагола в българския език и

продължителните и перфектните времена в английския. Това е значим извод, който пряко засяга теоретичното класифициране на тези граматични признаци, отнасяни нерядко към едно по-широко семантично поле.

Конкретните данни от обработката на използвания корпус говорят, че в 70% от случаите темпоралната форма в българския глагол еднозначно може да предопредели темпоралната форма в английския език (с. 172), а това в от съществено значение за машинния превод. Важни са анализите на докторанта за причините за несъответствието между формално-семантичния анализ и емпиричните данни от Корпуса, защото засягат най-важните отлики между семантиката и функционирането в текста на темпоралните системи в двата езика.

Направените наблюдения върху конкретните паралели в Корпуса и изнесените статистически данни са особено ценни доказателства, които верифицират изследователските хипотези и превръщат изводите в доказани научни факти.

6. Преценка на публикациите по дисертационния труд: брой; характер на изданията (международни, национални), в които са отпечатани. Какво е отражението им в науката – използване и цитиране от други автори.

По темата на дисертационния труд са направени 7 научни публикации. Две от тях са публикувани в сборници от един от най-значимите международни научни форуми по компютърна лингвистика у нас. Тези трудове, както и публикуваните статии в сп. „Български език“, сп. „Съпоставително езикознание“, в електронното списание *Littera et Lingua* и в сборника от Международната юбилейна конференция на ИБЕ от 2017 г. са достатъчно доказателство, че научната общност е имала възможност да се запознае със значими научни търсения и постижения на дисертационния труд.

7. Авторефератът изготвен ли е съгласно изискванията, правилно ли отразява основните положения и научните приноси на дисертационния труд.

Авторефератът отговаря на структурата на дисертационния труд и отразява вярно постиженията му. Научните приноси представят адекватно положителните резултати от изследването.

8. Заключение с ясно становище да бъде ли присъдена научната степен, или не.

В резултат на анализа на научните и научно-приложните приноси на дисертационния труд и тяхната значимост за езиковедската наука и за компютърната лингвистика убедено предлагам на Тодор Димитров Лазаров да бъде присъдена научната и образователна степен „доктор“ за дисертационния труд „Българско-английски граматични паралели с оглед на машинния превод“ и гласувам положително за това решение.

24.04.2020 г.

Рецензент:

(проф. д-р Крaсимиpа Aлeкcoвa)