



РЕЦЕНЗИЯ

за дисертационния труд на Тодор Димитров Лазаров
задочен докторант към Секция по компютърна лингвистика, ИБЕ - БАН
на тема „Българско-английски граматични паралели с оглед на машинния превод“,
Професионално направление 2.1. Филология, Научна специалност Общо и сравнително
езикознание (Компютърна лингвистика)
научен ръководител проф. д-р С. Коева
от проф. д-р Радка Влахова-Руйкова, СУ „Св. Климент Охридски“

Данни за докторанта

Тодор Лазаров е завършил специалността Българска филология и Магистърска програма „Компютърна лингвистика. Интернет технологии в хуманитаристиката“ в СУ „Св. Климент Охридски“, ФСлФ. Приет е като редовен докторант по Общо и сравнително езикознание (Компютърна лингвистика), като формата на докторантурата по-късно е трансформирана в задочна. Докторантът притежава отлични дигитални умения, ползва 4 езика – английски, руски, немски, нидерландски.

По време на докторантурата Лазаров е участвал в Лятна школа по дигитална хуманитаристика и има две краткосрочни специализации – в Университета на Уълвърхемптън, Великобритания и в Университета в Осло, Кралство Норвегия. По време на специализациите е изнасял лекции върху проблематиката на дисертацията. Трябва да се подчертае, че участието на докторанта в обученията и специализации има пряко отражение в дисертационната разработка. Впечатление прави и участието му в научни проекти. Награден е за най-добър проект по Програма за насърчаване на млади учени през 2017 за ръководения от него проект.

Т. Лазаров е бил стажант към СКЛ – ИБЕ, БАН, бил е лектор в магистърски програми на НБУ, в момента работи в преводаческа агенция.

Съгласно Доклад за изпълнение на изискванията според Правилник за задочните докторанти в БАН докторантът е изпълнил всички изисквания. Трябва да се отбележи, че по раздели II. и III. – апробация на резултатите от изпълнението на научната програма и публикации по темата, докторантът е кумулирал двойно повече от изискуемия минимум кредити. Т. Лазаров е участвал с доклади на CLIB 2017, CLIB 2019, Международна юбилейна конференция на ИБЕ, както и на два международни форума, организирани от Дирекция „Писмени преводи“ на Европейската комисия (2018, 2019). Публикациите са в реферирани и индексирани издания – сп. „Български език“, сп. „Съпоставително езикознание“, както и в сборниците с материали от международните научни форуми.

Структура на текста

Изследването е структурирано в: *Обща характеристика на дисертационния труд*, където са представени актуалността и мотивацията на изследването, целите, обектът и предметът и методът на изследването; *Българско-английски паралели при граматичната категория време: Анализ на граматичните паралели в „Българско-английския паралелен корпус със съотнесени глаголни форми“: Практическо приложение и оценка на „Българско-английския паралелен корпус със съотнесени глаголни форми“*; *Заклучение*, което включва обобщение и приноси, *Библиография* и *Списък с публикациите по темата на дисертацията*. Дисертацията съдържа 197 с., библиографията включва 141 източника от български и чуждестранни учени, представителни за темата на изследването. В Приложение е достъпен създаденият от дисертанта корпус.

В уводната част са представени актуалността на проблематиката, мотивацията за избор на тема, формулирани са основните цели и задачи, посочен е основният подход. Още в тази уводна част Т. Лазаров се заявява като изследовател със солидни лингвистични познания и способност ясно да формулира целите и задачите на изследването. Темата е дисертабилна. Изследването има теоретичен и теоретично-приложен характер и е посветено на една от най-актуалните и динамично развиващи се области на компютърната лингвистика – машинния/ автоматичен превод.

Втора глава. *Българско-английски паралели при граматичната категория време*. В тази глава искам да отбележа на първо място лингвистичната осведоменост на докторанта и

зрелостта, с която са изведени основните проблеми при съпоставка на две несиметрични, на първо място неравнобройни, системи - 9 грамемии в български и 16 в английски.

Изследването на Т. Лазаров е принос към изследванията върху българския глагол – граматическите категории, които притежава българският глагол и граматическите значения, които изразява. Обстойно и задълбочено дисертантът представя различните подходи и концепции на български и чуждестранни езиковеди за темпоралната система на глагола в български и в английски, както и подходите при наличие на асиметрия. Аргументирано и въз основа на богат емпиричен материал, са разграничени формите, които изразяват морфологически сигнализираната категория „вид на действието“. На полето на българския глагол са се водили истински филологически „битки“ и обикновено изследователите на глаголните категории и техните значения се придържат към една или друга теория, респ. описателен модел. Докторантът е успял максимално детайлно да опише морфологично сигнализираните значения на грамемите, които са обединени в българската и в английската темпорална система. В труда се прави разграничаване на категорията *вид на действието* и *темпоралност*. Разгледана е категорията *вид на действието* във връзка с опозицията *результативност : нерезультативност/аукционност*, както и отношението *таксис : относителност*. Както за български, така и за английски в труда е направено цялостно описание на грамемите, обединени в темпоралната система. Представен е подробен анализ на словоформите, като всички те са богато илюстрирани.

В изследването си Т. Лазаров не само съпоставя различните описателни модели, а обосновава изтъква онези компоненти в значението, които могат да бъдат формализирани и да се включат в съпоставителния анализ с оглед на машинния превод. Разграничени са морфологично сигнализираните значения от синтактичните функционални съответствия, аналитичните форми и значенията, които се изразяват по еквиполентен, лексико-граматически път. В Табл. 4 (с. 78) прегледно са представени значенията на грамемите в двата езика, а в Табл. 5 (с. 79) са съотнесени отделните словоформи. Прави впечатление, че най-голяма – 12:1/2, е разликата при формите за *минало неопределена, бъдеще предварително* и *бъдеще предварително време в миналото*. Разликата се дължи основно на наличието само на I форма за лице и число при съответните темпорални форми в английски.

Дисертантът е разгледал типологията на езиците с оглед на частите на речта и мястото на глагола в системата. Много добре е направено разграничението между граматическо (семантично по природа) и наративно време. Времето е разгледало като хиперкатегория, чиито значения се реализират в йерархична система от признаци. Т. Лазаров прави съпоставка на формообразуването в диахронен и в синхронен план и посочва основната отлика между двата езика по отношение на категорията *вид на действието* – морфологична категория за български и синтактична категория за английски. Много добре е представена и ролята на словореда за идентифициране на подлога в английски. Тук искам да отбележа и отношението, което дисертантът взема към въпроса дали личните местоимения в позиция на подлог следва да се включат в глаголната парадигма като изразители на значението *лице* в английски. Независимо от несиметричността на формите, които изразяват това значение в български и английски, Т. Лазаров правилно приема, че няма основание за такова решение.

Съпоставката на значенията на темпоралните форми в български и в английски, която се основава на модела на Райхенбах, доразвит за българския глагол от Ницолова, която въвежда интервали, които от своя страна могат да бъдат отворени/затворени/отворени само от една страна – лява или дясна, представена прегледно в табл. 5., е ясна, убедителна и съотнесима с описанията на значенията на темпоралните форми и у другите автори – Куцаров, Герджиков, Пашов, за български и на описанията на авторите, изследващи английския глагол. Основните въпроси, които формулира дисертантът с оглед на задачите, които си поставя, са в две насоки: относно трансфера на граматикализираната информация от темпоралните глаголни форми – начините, по които тази информация се предава при превод, и степента на трансфер между двата езика. Основният въпрос е, може ли да се очаква граматичните категории, изразяващи сходни значения в двата езика, да съответстват напълно при превод и съответно категориите, които нямат съответните корелати, по какъв начин трансферират значението си. Другата група въпроси са относно типовете формообразуване и трансфера на информацията, носена от съответните морфемии – дали и в каква степен наличието на определени морфемии, натоварени със съответното граматично значение в български, предполага наличието на съответните морфемии в английски.

Трета глава *Създаване на българско-английски паралелен корпус със съотнесени глаголни форми* се спира първо на историята на машинния превод, отново изчерпателно и компетентно, и на езиковите корпуси като ресурси за машинен превод. Като се опира на трансферните правила като вид граматика, дисертантът обосновава корпусно базирания подход и статистическото езиково моделиране като най-подходящи за целите на изследването (с. 105-106). „Резултатите от работата по създаването на лингвистично обогатен статистически модел за автоматичен превод ще бъдат основно в две насоки: теоретична – изясняване на зависимостите при превод на български глаголни форми на английски, и практическа – обогатяване на статистически преводни модели за автоматичен превод с лингвистична информация за преводното съотнасяне на глаголни форми, изразяващи различни категории“ (с. 106).

За изпълнение на задачите, които си е поставил, дисертантът е създал двуезиков паралелен корпус със съотнесени темпорални форми в български и в английски, който е илюстративен (за поставената си цел). Създаденият от Т. Лазаров корпус съдържа 5070 аотирани и съотнесени глаголни форми и техните преводни съответствия. В крайния си вид „Българско-английският паралелен корпус със съотнесени глаголни форми“ се състои от 84% текстове от „Българско-английския паралелен корпус със съотнесени изречения“, 15,5% текстове от интернет и 0,5% от изречения, извлечени от примери от „Българска граматика“.

Съотнесените глаголни форми и техните преводни съответствия са представени прегледно и илюстрирани в поредица от таблици за всяко глаголно време. Във Втора глава дисертантът е обърнал внимание на наличието на конкорданс в английски и отсъствието му в български. Зависимата употреба на форми за сегашно време в подчинено изречение в български се обуславя от събитието в главното, което играе ролята на момент на референтността. На този етап дисертантът не си поставя задача да се извлече статистика колко от примерите са употреби в просто и колко в сложно изречение. Това може да бъде предмет на бъдещо изследване, тъй като много от примерите показват, че информацията се идентифицира на тези равнища. Статистическите данни, изведени от дисертанта са надеждна основа както на машинно подпомогнат превод, така и за теоретични изследвания върху глагола в български и в английски. Интересен паралел може да се открие и при формите за перфект – 6,05% в български и 6,40% инфинитив в английски.

В поредица от прегледни таблици са представени и анализирани изразяваните от темпоралните форми граматически значения. Те представят описаните равнища на анотация и съответната статистика. Данните за двата езика са съотнесени по същите показатели. Дисертантът мотивирано е избрал и критериите за съпоставка- въз основа на маркирани и немаркирани форми, респ. с основно значение, въз основа на описанието, възприето от Куцаров, и на прагматико-функционалните съответствия, основаващи се на модела на Райхенбах. В този смисъл Т. Лазаров се спира и на категорията вид във връзка с продължителните времена в английски.

Докторантът правилно заключава, че „Данните от Корпуса показват, че в значителна степен лексикалният контекст и синтактичните особености на езиците, както и прагматичната употреба предопределят използването на темпоралните форми. Граматикализираната информация се запазва в значителна степен, но се пренася също така и чрез допълнителни езикови елементи – лексикален материал или контекстна съгласуваност между отделните темпорални форми в рамките на текста. Статистическите данни също показват, че изследването на морфологичните кореспонденции между темпоралните системи в български и английски и преводните реализации чрез „Българско-английския паралелен корпус със съотнесени глаголни форми“ е възможно и ползотворно.“ (с. 175)

Задача на корпуса, както определя дисертантът, е да се създадат обогатени преводни паметни. Надежден показател е, че 85% от първото предложение съвпада с избора на преводача. Важен резултат е, че няма избор, който да не съвпада с нито едно предложение (с. 185). В дисертационното изследване е представена обективна количествена и дистрибутивна информация за глаголните системи на български и английски, подходяща за различен тип научен анализ извън обхвата на проучването. Осигурени са надеждни лингвистични данни за създаването на различни езикови приложения за компютърна обработка на естествения език като статистически модел за анотация и статистически модел за превод на глаголните форми. Трябва да се изтъкне, че лингвистичните данни от „Българско-английския паралелен корпус със съотнесени глаголни форми“ са интегрирани в софтуер за компютърно подпомогнат превод с цел практическото му приложение като помощно средство при превод.

Езикът на текста е ясен и точен. Авторефератът коректно отразява съдържанието на дисертацията, изводите и приносите на изследването.

В заключени смятам, че дисертантът е изпълнил задълженията си съгласно Правилника за задочни докторанти на БАН. Представеният за публична защита текст на тема „Българско-английски граматични паралели с оглед на машинния превод“ показва задълбочена лингвистична подготовка, изследователски талант, умение да се създава научен продукт. Аprobацията и публикациите на научните резултати са на авторитетни научни форуми и в реферирани издания. Резултатите са убедителни и добре аргументирани. Налице са всички основания – формални и по същество убедено да подкрепя с положителен вот да бъде присъдена образователната и научна степен ДОКТОР на Тодор Димитров Лазаров по Професионално направление 2.1. Филология, Научна специалност Общо и сравнително езиковедие (Компютърна лингвистика).

18. 04. 2020

Подпис:

(проф. Р. Влахова-Руйкова)