



## СТАНОВИЩЕ

От члена на НЖ Максим Иванов Стаменов, д.ф.н., проф. в Института за български език „Проф. Любомир Андрейчин“ към БАН

за дисертационен труд за придобиване на (образователната и) научна степен .доктор.

Автор на дисертационния труд: **Тодор Димитров Лазаров**, докторант в Института за български език „Проф. Любомир Андрейчин“ при БАН.

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

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

Проблемът, на който е посветена предлаганата дисертация, е твърде актуален в контекста на най-новите насоки на развитие на компютърната лингвистика. Трудът е ориентиран към изследването на българско-английските граматични паралели с оглед на машинния превод и, по-конкретно обектът на изследване е ограничен до анализ и моделиране на темпоралните системи на двата езика и изразяването на темпоралните отношения в български и английски и начините за предаване на граматичното значение на глаголните форми от български към английски. От една страна, изследването на тази проблематика чрез подходите и методологията на машинния превод е мотивирано от възможностите, които научното поле предлага за систематичен и формален анализ на комплексни езикови явления, каквито са и темпоралните системи и глаголното формообразуване. От друга страна, слабата разработеност на проблема и липсата на достатъчно езикови корпуси, които имат анотация и информация за съотнасянето на глаголните форми, предполага теоретико-приложния аспект на изследването.

Специално трябва да се изтъкне съчетаването в дисертацията на теоретичния и научно-приложния аспект в нея. Основната цел на дисертацията е да се направи съпоставка на теоретично равнище между семантиката на темпоралните значения в двата езика, което ще подпомогне теоретичното обосноваване на възможните преводни еквиваленти, и статистически анализ на данни, извлечени от паралелен корпус със съотнесени глаголни форми между български и английски. Резултатите от изследването са ориентирани с оглед на прилагането им за усъвършенстването на работата на системите за компютърно подпомогнат превод, както и при подбора на едоезикови и паралелни данни за обучение на системите за машинен превод.

С оглед на тази ориентация и са формулирани целите на работата: да бъдат разгледани българско-английските граматични паралели при превод на темпоралните форми от български на английски и да бъдат представени статистически данни за граматикализираната информация от темпоралните форми при превод от български на английски.

### 2. Ориентация на дисертанта за състоянието на проблема и научната литература

Авторът показва добро познаване на литературата по темата на дисертацията. В библиографията към дисертацията са посочени общо 149 публикации на наши и чуждестранни учени, с които дисертантът е работил при извършването на дейностите, свързани с извършването на изследването и написването на дисертационния труд.

### **3. Характер на избраната методика**

Методът е този на изследване на възможностите на формално-семантичното описание на свойствата на езика и използване на този тип описание за приложни цели. Разглеждането на хиперкатегорията време в български и английски предоставя възможността за паралелен формално-семантичен анализ на значенията на темпоралните категории. Проблемът за граматичните паралели между темпоралните системи на двата езика се разглежда чрез методите на семантичния анализ, корпусния анализ и статистическото моделиране. Граматичните паралели между български и английски с оглед на формално-семантичното описание на сходствата и разликите на темпоралните им системи се подлага на емпирична верификация на базата на статистически данни, извлечени от „Българско-английския паралелен корпус със съотнесени глаголни форми“. За целта корпусният анализ е съчетан със статистическо моделиране.

### **4. Кратка аналитична характеристика на научните и/или научно-приложните приноси на дисертационния труд**

Потвърждаваме наличието на следните приноси в труда. На първо място това се отнася до представянето на единно в теоретично отношение описание на темпоралните системи в български и английски както поотделно, така и в съпоставка. В съпоставителен план е изследван механизмът на конструиране на граматичното значение на темпоралните глаголни форми в български и английски и са формулирани основните закономерности при взаимодействието между темпоралните морфологични категории в двата езика и концептуалната структура на хиперкатегорията време.

В дисертацията е конструиран и аотиран Българско-английски паралелен корпус със съотнесени глаголни форми. Корпусът представлява езиков ресурс, който предоставя възможности за изследването на граматическите паралели между български и английски език. Новото при него е, че той изпълнява функцията на средство за формалното описание на граматичните паралели между темпоралните системи на български и английски и източник за извличането на статистически модели за отношенията на преводните реализации на глаголните форми в двата езика. При този начин на конструкция и употреба езиковите данни от Корпуса могат да послужат за обогатяване на преводни модели за нуждите на машинния превод, както и за теоретични изследвания. Представената процедура по структурирането и аотирането на Корпуса може по-нататък да се използва при разработването на подобни лингвистични ресурси за изследване на сходни езикови явления извън обхвата на обсъжданото изследване.

Съществени са обобщенията на базата на количествена и дистрибутивна информация на глаголните системи на български и английски, подходящи за различни типове научен анализ, които са извън обхвата на проучването. По този начин са осигурени надеждни лингвистични данни за създаването на алтернативни езикови приложения за компютърна обработка на естествения език като статистически модел за аотация и статистически модел за превод на глаголните форми.

Допълнителен принос е интегрирането на лингвистичните данни от Българско-английския паралелен корпус със съотнесени глаголни форми в софтуер за компютърно подпомогнат превод с цел практическото му приложение като помощно средство при превод. По този начин статистическите данни могат както да бъдат използвани при процеса на превод, така и да бъдат коригирани и обогатени с допълнителен езиков материал.

Не може обаче да се приеме за приносен посоченият от кандидата труд, положен за представяне на процеса по подбор на ресурсите, за предварителната обработка, анотацията и за използване на софтуерни приложения за работата по корпуса. Ако става въпрос за формирането на самата процедура по структурирането и аотирането на Корпуса и възможностите тя да бъде използвана за други цели, тя вече беше спомената при обсъждането на приноса на Българско-английския паралелен корпус със съотнесени глаголни форми.

#### **5. Преценка на публикациите по дисертационния труд: брой, характер на изданията, в които са отпечатани, цитирания**

Дисертантът е публикувал по темата на дисертацията общо седем публикации. Всички те са под негово авторство. Три от неговите публикации са в реферирани издания – две в сп. „Български език“ и една в сп. „Съпоставително езикознание“. Има и една публикация в ел. списание „Littera et Lingua“. Две от тях са на английски език и са публикувани в материалите от международни конференции по компютърна лингвистика, проведени в България (CLIB 2016 CLIB 2018). Една е публикувана в материалите от юбилейната конференция на ИБЕ, проведена през 2017 г. Липсват сведенията за статии на дисертанта да са били цитирани от други автори. Описаните публикации надхвърлят значително минималните изисквания за придобиване на научната степен „доктор по филология“.

#### **6. Критични бележки**

При представянето на особеностите на българската и английската система на глагола се включва и история на въпроса, което не винаги изглежда целесъобразно с оглед на избраната тема. Така например при обсъждането на английската система на глагола се прави преглед на възгледи на лингвисти от XVI век насам (с. 44, 46-48).

Теоретичната насоченост, колкото и да е похвална в рамките на изследване, ориентирано в руслото на компютърната лингвистика, има в някои случаи и обратна страна – спекулативни твърдения: „Съвместното присъствие на всички тези значения в текста изисква формален анализ на инвариантното значение, което обосновава разглеждането на позицията на формата в системата на езика на екстралингвистично ниво (Чергова 2014: 187)“ (с. 52).

Проблематично е твърдението, че „Когато действието в подчиненото изречение е бъдещо по отношение на действието в главното, употребява се бъдеще време в миналото (Ранкова и Иванова 1980: 226). То действа при структуриране на сложни изречения и последователни изказвания и е по-скоро синтактична особеност на изграждане на текст.“ (с. 61). Sequence of tenses е приложимо на равнище изречение, а не на текст. Нетривиален може да изглежда въпросът доколко в разговорната реч едно изречение може да се оформи като две изказвания, които се допълват. И в този случай според мен печели условието за добра оформеност с оглед на изреченския синтаксис.

7. **Авторефератът правилно ли отразява съдържанието на дисертационния труд**

Авторефератът правилно отразява основните положения и научните приноси на дисертационния труд. Той отговаря на съдържанието му.

8. **Заклучение**

Имайки предвид гореизложеното, предлагам на Научното жури да **присъди** на дисертанта **Тодор Димитров Лазаров** научната и образователна степен „доктор” по професионално направление 2.1 Филология, научна специалност Общо и сравнително езикознание (компютърна лингвистика) за дисертацията му „**Българско-английски граматични паралели с оглед на машинния превод**“.

Дата 25.04.2020 г.  
(подпис)

Изготвил становището:

