

СЕКЦИЯ ПО КОМПЮТЪРНА ЛИНГВИСТИКА
ИНСТИТУТ ЗА БЪЛГАРСКИ ЕЗИК „ПРОФ. ЛЮБОМИР АНДРЕЙЧИН“
БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ



Автореферат на дисертация за присъждане на образователната и научната
степен „доктор“

БЪЛГАРСКО-АНГЛИЙСКИ ГРАМАТИЧНИ
ПАРАЛЕЛИ С ОГЛЕД НА МАШИНИЯ ПРЕВОД

Професионално направление
2.1. Филология

Научна специалност
Общо и сравнително езикознание
(Компютърна лингвистика)

Автор: Тодор Лазаров
Научен ръководител: проф. д-р Светла Коева

Състав на научното жури:
проф. дфн Максим Стаменов
проф. д-р Йовка Тишева
проф. д-р Красимира Алексова
проф. д-р Радка Влахова
доц. д-р Екатерина Търпоманова

Резервни членове на научното жури:
проф. д-р Велка Попова
проф. д-р Диана Благоева

Дисертационният труд е обсъден на разширено заседание на Секцията по компютърна лингвистика на Института за български език „Проф. Л. Андрейчин“ при БАН на 27.01.2020 г. и е насочен за защита с Протокол № 2/30.01.2020 г. на Научния съвет на Института за български език „Проф. Л. Андрейчин“ при БАН.

Дисертацията съдържа 197 страници. Трудът се състои от увод, четири глави, заключение, библиография и приложение.

Съдържание

I. Обща характеристика на дисертационния труд.....	3
I.1. Актуалност и мотивация на изследването	3
I.2. Цели на изследването	4
I.3. Обект и предмет на изследването	8
I.4. Метод на изследване	9
II. Българско-английски паралели при граматичната категория време.....	11
II.1. Времето като глаголна категория – концепции и дискуссионни въпроси ...	11
II.2. Характеристика на глаголите в български	13
II.2.1. Морфологичните категории лице, число и род при глаголите в български.....	14
II.2.2. Особености на българската темпорална система	14
II.2.3. Семантична характеристика на българската темпорална система.....	17
II.3. Характеристика на глаголите в английски.....	18
II.3.1. Категориите лице и число на глаголите в английски	19
II.3.2. Особености на английската темпорална система.....	20
II.3.3. Семантична характеристика на английската темпорална система.....	21
II.4. Сравнителен преглед на характеристиките на глаголните системи и начините за образуване на темпоралните форми в български и английски	23
III. Създаване на „Българско-английския паралелен корпус със съотнесени глаголни форми“	29
III.1. Машинен превод – цели и историческо развитие	29
III.2. Езиковите корпуси като ресурс за машинен превод	30
III.3. Мотивация за избора на корпусния подход при изследването на българско-английските граматични паралели	31
III.4. „Българско-английския паралелен корпус със съотнесени глаголни форми“	32
III.4.1. Подбор на езиковите ресурси, включени в Корпуса	32
III.4.2. Структура на „Българско-английския паралелен корпус със съотнесени глаголни форми“	34
III.4.3. Анотационни конвенции при създаването на „Българско-английския паралелен корпус със съотнесени глаголни форми“	35
III.5. Характеристика на „Българско-английския паралелен корпус със съотнесени глаголни форми“	37
IV. Анализ на граматичните паралели в „Българско-английския паралелен корпус със съотнесени глаголни форми“	39

IV.1. „Българско-английският паралелен корпус със съотнесени глаголни форми“ в контекста на изследването.....	39
IV.2. Статистически обусловен анализ на данните за българските темпорални форми в Корпуса	40
IV.3. Статистически обусловен анализ на данните за английските темпорални форми в Корпуса	40
IV.4. Статистически обусловен анализ на превода на темпоралните форми от български на английски	41
IV.5. Обобщен анализ на превода на темпоралните морфологични категории от български на английски	47
IV.6. Изводи за българско-английските граматични паралели на темпорални форми	51
V. Практическо приложение и оценка на „Българско-английския паралелен корпус със съотнесени глаголни форми“	53
Заключение	55
V.1. Обобщение	55
V.2. Приноси на изследването.....	57
Библиография.....	59
Списък с публикациите по темата на дисертацията	61

I. Обща характеристика на дисертационния труд

I.1. Актуалност и мотивация на изследването

Дигиталната обработка на езикови данни през последните години става все популярна и се използва все по-широко не само в областта на лингвистиката, но и в редица свързани области. Една от причините за нарастващата ѝ популярност е, че позволява на изследователите да обработват много големи количества текстови данни и да установяват различни текстови характеристики, общи черти, модели, които може да са трудни или дори невъзможни за установяване чрез класическите методи на анализ. Свързана тенденция, също доста силна и все по-популярна, е създаването на различни езикови корпуси, включващи различни видове езикови данни, за целите на компютърната обработка на езика и създаването на функционални езикови модели.

Актуалността на машинния превод през последните десетилетия е продиктувана от нуждата в глобален аспект от бърз достъп до различна по вид, обем и език информация. Основните преимущества на системите за машинен превод са гарантирането на непротиворечивост на превода, подобрена производителност и намалени разходи. Съществен недостатък на системите за машинен превод е, че с малки изключения са разработени и достъпни предимно за по-разпространени езици като английски, френски, немски, испански, руски, арабски и други.

Дисертацията е насочена към изследване на българско-английските граматични паралели с оглед на машинния превод. Обектът на изследване е ограничен до темпоралните системи на двата езика: изразяването на темпоралните отношения в български и английски и начините за трансфер на граматичното значение на темпоралните форми от български към английски. От една страна, изследването на тази проблематика във връзка с подходите и методите, използвани при машинния превод, е мотивирано от възможностите, които научното поле предлага за систематичен и формален анализ на комплексни езикови явления, каквито са темпоралните системи на двата езика. От друга страна, липсата на паралелни езикови ресурси, които съдържат информация за съотнасянето на глаголните форми при превод от български на английски език, предполага теоретико-приложния аспект на изследването.

I.2. Цели на изследването

Основната цел на дисертацията е да се представи единно теоретично описание на темпоралните системи в български и английски и да се проверят теоретично изведените хипотези за преводните еквиваленти посредством статистически анализ на данните от паралелен корпус със съотнесени темпорални форми между български и английски: „Българско-английския паралелен корпус със съотнесени глаголни форми“.

Резултатите от изследването могат да бъдат приложени при усъвършенстването на работата на системи за компютърно подпомогнат превод, както и при подбора на едоезикови и паралелни данни за обучение на системи за машинен превод.

За целите на дисертацията се въвеждат следните терминологични уточнения: с термина граматична категория време се означава традиционната деветчленна система на времената в български и шестнадесетчленната система на времената в английски (и инфинитив). Деветте времена в български са: сегашно време, минало свършено време, минало несвършено време, минало неопределено време, минало предварително време, бъдеще време, бъдеще предварително време, бъдеще време в миналото и бъдеще предварително време в миналото, а шестнадесетте времена в английски са: сегашно просто време (Present Simple Tense), сегашно продължително време (Present Continuous Tense), сегашно перфектно време (Present Perfect Tense), сегашно перфектно продължително време (Present Perfect Continuous Tense), минало просто време (Past Simple Tense), минало продължително време (Past Continuous Tense), минало перфектно време (Past Perfect Tense), минало перфектно продължително време (Past Perfect Continuous Tense), бъдеще просто време (Future Simple Tense), бъдеще продължително време (Future Continuous Tense), бъдеще перфектно време (Future Perfect Tense), бъдеще перфектно продължително време (Future Perfect Continuous Tense), бъдеще просто време в миналото (Future Simple Tense in the Past), бъдеще продължително време в миналото (Future Continuous Tense in the Past), бъдеще перфектно време в миналото (Future Perfect Tense in the Past) и бъдеще перфектно продължително време в миналото (Future Perfect Continuous Tense in the Past).

Приема се, че посредством граматичната категория време (още хиперкатегорията време) се изразяват различни (темпорални) отношения и тя може да се разглежда като съвкупност от три морфологични категории в български: време, вид на действието и

таксис, и две морфологични категории в английски: време и вид на действието. Морфологичната категория **време** (наричана още същинска категория време) **в български** „се основава на отношението на действието към изказването“ (Куцаров 2007: 244), което отношение може да изразява „едновременност или разновременност“ (неедновременност). Морфологичната категория **време в английски език** „изразява едновременност или неедновременност на действието спрямо момента на говоренето“ (Кабакчиев 2003: 132). Морфологичната категория **вид на действието** (результативност) **в български** изразява „отношение между действие и резултат от действие в един и същ ориентационен момент“ (Куцаров 2007: 249), който може да бъде сегашен, минал, бъдещ или бъдещ спрямо минал. Морфологичната категорията **вид на действието (вид) в английски** изразява отношението между „продължителността или завършеността на действието към [...] ориентационен момент“ (Валейка и Буйткийн 2003: 84), сегашен, минал, бъдещ или бъдещ спрямо минал. Необходимо е да бъде уточнено, че категорията вид на действието в български и категорията вид на действието в английски представляват две различни категории с различна семантика и категориална структура, въпреки съпадението на термините, които са възприети. Морфологичната категория **таксис в български** изразява „отношенията на действията или резултатите от действия, ориентирани към миналия ориентационен момент (минало време или друг сигнализатор), към действията, неориентирани към миналия ориентационен момент“ (Куцаров 2007: 255).

Представените дефиниции са възприети като най-ясно дефиниращи значението на съответните морфологични категории за български и английски. Тъй като нито един от авторите не дава дефинициите във връзка със съпоставката на темпоралните системи на двата езика, не може да се говори за обобщения, които имат типологичен характер.

С термина **темпорални отношения** се означават отношенията, изразявани от трите морфологични категории: време, вид на действието и таксис в български, и двете морфологични категории: време и вид на действието в английски. С термина **темпорална система** се означава системата от морфологичните категории време, вид на действието и таксис в български и време и вид на действието в английски. С термина **темпорални форми** се означават глаголните форми в двата езика, които са част от съответните темпорални системи.

Основната цел на дисертацията се постига последователно посредством целите и задачите, чието изпълнение е описано в отделните глави на дисертацията.

Целта на Втора глава „Българско-английски паралели при граматичната категория време“ е да се изгради единна концепция за изследване на темпоралните форми в двата езика, да се представи формално-семантично описание на темпоралните значения както поотделно в двата езика, така и съпоставителен план и да се създаде хипотеза за възможните съответствия на семантично равнище. Формално-семантичното описание на темпоралните значения в двата езика ще бъде използвано при изграждането на хипотеза за българско-английските граматични паралели на равнище темпорални форми.

Конкретните задачи за постигането на поставената цел във втората глава от дисертацията са:

1. Да се разгледат и съпоставят различни схващания за морфологичните категории време, вид на действието и таксис в български и време и вид на действието в английски.
2. Да се създаде единна за целите на настоящото изследване концепция за описание на темпоралните системи в български и английски език, морфологичните категории, които ги изграждат, и отношенията между членовете им.
3. Да се направи формално-семантично описание на темпоралните значения в двата езика.
4. Да се обособят граматичните паралели между формално-семантичните описания на темпоралните значения в двата езика, на основата на които ще бъде създадена хипотеза за преводните съответствия и семантичния трансфер на граматична информация при превод.

Целта, към която е ориентирано изследването в Трета глава „Създаване на „Българско-английски паралелен корпус със съотнесени глаголни форми“, е създаването на лингвистично анотиран паралелен корпус, съдържащ информация за преводното съотнасяне на българските темпорални форми с английските им съответствия. За осъществяването на поставената цел е необходимо да бъдат изпълнени следните задачи:

1. Анализ на основните характеристики на машинния превод и изискванията към езиковите ресурси, които се използват за машинен превод.
2. Преглед на съществуващите паралелни ресурси за български и английски с цел подбор на подходящите за целите на изследването ресурси.

3. Формулиране на критерии за изграждането на „Българско-английския паралелен корпус със съотнесени глаголни форми“.

4. Формулиране на анотационни конвенции за обогатяването на „Българско-английския паралелен корпус със съотнесени глаголни форми“ с лингвистична информация за съотнесените глаголни форми.

5. Практическа работа за създаване и анотиране на „Българско-английския паралелен корпус със съотнесени глаголни форми“.

6. Описание на основните характеристики на „Българско-английския паралелен корпус със съотнесени глаголни форми“.

Основната цел на изследването в Четвърта глава е да се представи статистически обусловен анализ на темпоралните форми в български и английски както поотделно, така и в съпоставителен план. Анализът е в няколко направления: по отношение на данните за българските и английските темпорални форми поотделно; по отношение на трансфера на информация за морфологичните категории време, вид на действието и таксис в български към английските морфологични категории време и вид на действието, както и по отношение на трансфера на информацията за морфологичните категории лице, число и род и лексикално-граматичната категория вид на глагола в български при превод на английски език. Конкретните задачи за постигането на поставената цел в четвъртата глава от дисертацията са:

1. Извличане на данни и статистически обусловен анализ на българските темпорални форми, разделени по темпорални категории и тематични области в Корпуса.

2. Извличане на данни и статистически обусловен анализ на английските темпорални форми, разделени по темпорални категории и тематични области в Корпуса.

3. Статистически обусловен анализ на превода на темпоралните форми от български на английски.

4. Формулиране на изводи за българско-английските граматични паралели на темпоралните форми.

5. Представяне на основните резултати от анализа на данните в „Българско-английския паралелен корпус със съотнесени глаголни форми“ в контекста на теоретичното описание на българската и английската темпорална система.

Друга основна цел на изследването е приложение на данните от Корпуса за създаването на преводни модели, обогатени с лингвистична информация за съотнасянето на български и английски темпорални форми. По-конкретно, практическо приложение на „Българско-английския паралелен корпус със съотнесени глаголни форми“ в система за компютърно подпомогнат превод. За изпълнението на тази цел се реализират следните задачи, представени в Пета глава „Практическо приложение и оценка на „Българско-английския паралелен корпус със съотнесени глаголни форми“:

1. Създаване на „терминологична“ база от данни, съдържаща информация от „Българско-английския паралелен корпус със съотнесени глаголни форми“.
2. Провеждане на експеримент с терминологичната база от данни при компютърно подпомогнат превод.
3. Оценка на резултатите от експеримента.

В заключение, целта на дисертацията може да бъде представена като съвкупност от поставените цели по отношение на няколко концептуално отделни части. От една страна, дефинирането на единна концепция за описание на темпоралните системи и формообразуването в български и английски е необходима предпоставка за съпоставянето на двата езика на теоретично равнище. От друга страна, създаването на хипотеза, базирана на теоретичното описание на темпоралните системи на двата изследвани езика, позволява да се направи оценка на извлечените статистически езикови данни от „Българско-английския паралелен корпус със съотнесени глаголни форми“.

I.3. Обект и предмет на изследването

Обект на изследването са съвременните темпорални системи в български и английски, начините на изразяване на хиперкатегорията време в двата езика, семантиката и трансфера на морфологичните категории време, вид на действие и таксис в български към морфологичните категории време и вид на действието в английски. Предмет на изследване са конкретните граматични паралели между темпоралните отношения в двата езика, засвидетелствани в паралелен корпус, създаден за нуждите на проучването.

Предназначението на „Българско-английския паралелен корпус със съотнесени глаголни форми“ е да представя изчерпателна и достоверна езикова информация, приложима за създаването на статистически модел за превод на глаголните форми.

Необходимо е да се направи уточнението, че при създаването на „Българско-английския паралелен корпус със съотнесени глаголни форми“ не са отразени евиденциалността, наклонението, залогът, положителният или отрицателният статус. Дефинирането на езикова посока (от български към английски) е продиктувано от комплексността на българското глаголно формообразуване в сравнение с английското.

I.4. Метод на изследване

В последните години се откроява подчертана тенденция към формално-семантично описание на свойствата на езика и към приложение на това описание за практически цели. Разглеждането на хиперкатегорията време в двата езика предоставя възможността за паралелен формално-семантичен анализ на значенията на темпоралните категории.

Трябва да се отбележи, че изследването във формален план на особеностите на формообразуването в български и английски (както отделно за двата езика, така и в съпоставителен план) не е новост; съществуват изследвания, акцентиращи върху по-обща или върху по-специфични проблеми; известни са и опити за представянето на формални граматични паралели за превод между български и английски – както за глаголите, така и за останалите части на речта (Атанасова и др. 1963, Стамболиева 2015). В дисертация проблемът за българско-английските граматични паралели между темпоралните системи в двата езика се разглежда чрез методите на формално-семантичния анализ, корпусния анализ и статистическото моделиране.

Избраните методи за анализ на българско-английските граматични паралели на равнището на темпоралните системи на двата езика са тясно обвързани с компютърната лингвистика и по-специално с машинния превод. Изследването е повлияно от съвременните тенденции за системен анализ на езикови феномени чрез прилагане на методи от компютърната лингвистика и статистическото моделиране (Паскалева 1987: 32). Актуалността на статистическото моделиране предполага интегриране на съвременните тенденции за изследване на езика и класическите теоретични изследвания.

Грамматичните паралели между български и английски, и по-специално формално-семантичното описание на различията и сходствата на темпоралните системи в двата езика, се подлага на емпирична верификация на базата на статистически данни,

извлечени от „Българско-английския паралелен корпус със съотнесени глаголни форми“. За целта корпусният анализ се съчетава със статистическото моделиране.

При машинния превод се разчита на представителни паралелни езикови корпуси с достатъчен обем, които представят лингвистичните феномени в техните различни проявления. За създаването на статистически езиков модел за превод е необходимо да се знае каква е вероятността дадена езикова единица (дума, фраза) в целевия език да бъде превод на езикова единица f от изходния – $p(e|f)$. За целта са необходими алгоритми, които да обработват информацията, и езикови ресурси, от които да се извлекат лингвистичните данни. Както е известно, „корпусите [...] представляват надежден източник за наблюдение, анализ и изводи (подкрепени от обективни количествени и дистрибутивни данни) за [...] различни езикови явления (за употребата им в различни стилове, жанрове и тематични области), както и за автоматично извличане на езикови данни, езикови отношения и модели“ (Коева 2014: 49).

Граматичните паралели между български и английски и по-специално формалното описание на различията и сходствата на темпоралните системи в двата езика се основава на анализа на данните от „Българско-английския паралелен корпус със съотнесени глаголни форми“, създаден за нуждите на изследването.

Достоверността на резултатите е проверена експериментално с помощта на софтуер за компютърно подпомогнат превод.

II. Българско-английски паралели при граматичната категория време

Във Втора глава на дисертацията е представено теоретично описание на темпоралните системи в български и английски и на формално-семантичните характеристики на двете системи.

II.1. Времето като глаголна категория – концепции и дискуссионни въпроси

Описанието на времето като феномен отдавна е обект на изследване както във философията, така и в лингвистиката. Аристотеловата традиция представя времето като „непрекъсната линия, която тече от ляво надясно, т.е. от миналото към бъдещето“ (Манинг и др. 2013: 233). В лингвистиката времето се разглежда като граматична категория, която „граматикализира протичането на събития или състояния във времето чрез локализиране на възможните събития“ (Комри 1985: 9).

Както в българското езикознание, така и в изследванията за английски език категорията време винаги е била дискуссионна по отношение на обхвата си: инвентара от граматични значения и кодирането на граматична информация. В наши дни е възприето схващането за т. нар. „хиперкатегории“, съставени от взаимосвързани категории със собствени граматични парадигми, които са с кумулативно значение по отношение на формообразуването (Чергова 2014: 195). Категорията време се отъждествява със семантичния континуум „отношение на изказаното действие към момента на говорене“ (Чергова 2014: 203), но в рамките на това значение може да се обособят подзначения за пряко и непряко отношение на действието към момента на говорене, което формира две подкатегории в темпоралността: глаголно време и таксис (Чергова 2012: 18 – 19).

Все по-продуктивни от функционална гледна точка се оказват похватите, които разглеждат глаголните категории като „сложни функционално-семантични структури с разнообразни езикови инструменти“ (Вейга 1991: 56) или хиперкатегории, при които „между крайните елементи на отношението, изразявано от съответната проста категория, се вмъкват един или повече елементи посредници“ (Герджиков 2000). „Сложните категории (хиперкатегориите) се състоят от синкатегории, които нямат равноправно положение в системата поради елементите, които изграждат значенията им (само крайни елементи, крайни елементи и елементи посредници и т.н.)“ (Алексова 2003).

От гледна точка на темпоралната логика са възможни три времеви отношения на момента на събитието спрямо момента на говорене (речевия акт): „преди момента на говорене“, „едновременно с момента на говорене“ и „след момента на говорене“ (Райхенбах 1947: 287). Граматичното време в много строг смисъл съдържа само две семантични стойности: едновременност и неедновременност (Райхенбах 1947: 291). Други темпорални отношения се формират с допълнителни категории като аспект или относителност. Разграничаването на аспекта и граматичното време не е еднотипно реализирано при различните езици, макар че те могат да се разглеждат като взаимозависими и обвързани (Хорнщайн 1990: 9). Аспектът в български е лексикално-граматична категория и формално остава извън рамките на хиперкатегорията време, докато в английски се изразява от категорията вид на действието (продължителност и перфективност), а в някои случаи и на лексикално равнище (посредством фразеологични глаголи).

За нуждите на формално представяне на темпоралната система на Х. Райхенбах ще се използва съкращението S, за да се означаи моментът (времето) на говорене. Времето на събитието, изразявано от глагола, ще бъде означено с E. Има две възможни отношения между S и E: отношение на последователност с две подразделения – E предхожда S и обратно; и отношение на едновременност – E и S се случват по едно и също време и S се включва в E.

Много важна характеристика на темпоралната система на Х. Райхенбах е, че тя се характеризира с три темпорални точки: вече дефинираните точки E и S и така наречената „отправна точка“ (R) (Райхенбах 1947: 293). Освен три темпорални точки (S, E и R) системата включва и два вида отношения между тях (едновременност и последователност). От решаващо значение е, че тези типове отношения могат да бъдат само между точките R и S или R и E, отношението между S и E е непряко и е изградено на базата на другите отношенията в системата (Клайн 1994: 54). Следователно отправната точка R може да се разглежда като свързващ възел в цялата система. Отношенията между S и R, от една страна, и E и R, от друга страна, според системата на Х. Райхенбах (Райхенбах 1947: 295) са представени в Таблица 1:

Отношение	Значение	Наименование	Отношение	Значение	Наименование
S, R	S е едновременно с R	Настояще	E, R	E е едновременно с R	Едновременност
S – R	S предхожда R	Бъдеще	E – R	E предхожда R	Предходност
R – S	R предхожда S	Минало	R – E	R предхожда E	Следходност

Таблица 1. Описание на отношенията между елементите в системата на Х. Райхенбах.

II.2. Характеристика на глаголите в български

Единствено в български в сравнение с останалите славянски езици глаголят запазва повечето най-важни особености от старобългарския език, но и претърпява допълнително развитие (Ницолова 2008: 224). Според морфологичните си характеристики глаголят попада в групата на изменяемите части на речта, а в глаголната парадигма се включват голям брой словоформи с множество граматични значения. Според синтактичната си функция и характеристики глаголят попада в групата на самостоятелните части, тъй като той е тази част от изречението, която изразява отношението между другите съставни части на изречението – той свързва аргументите в рамките на пропозицията (Ницолова 2008: 224).

Основното лексикално деление при глаголите на лексикално-граматични класове, което се основава на техните функционални особености, е на пълнозначни и спомагателни глаголи (Куцаров 2007: 108). Друго основно деление на глаголите е на преходни и непреходни според възможността за съчетаване с пряко допълнение, ако действието засяга пряко даден обект. Според възможността глаголите да изразяват граматичното значение лице, те се делят на лични и безлични (Куцаров 2007: 105).

Грамматическата категория вид на глагола „дава възможност протичането на едно действие да се представи по два начина: като завършено действие и като неприключил процес. Следователно видът е морфологическа категория, която показва начина на протичане на глаголното действие по отношение на признака цялостност (изчерпаност, завършеност, комплексност)“ (Гаравалова 2003: 186). Тази отличителна характеристика на глаголите в българския език е обект на множество научни дискусии и различни концепции. Възприема се, че делението на глаголите в българския език на глаголи от свършен вид и на глаголи от несвършен вид е в рамките на лексикално-граматичната категория вид на глагола (Куцаров 2007: 551). Според делението в тази категория глаголите се разпределят в зависимост от това дали действието е насочено към някакъв предел, или не, и съответно се означават като пределни и не пределни. „Само пределните глаголи са представени в свършен и несвършен вид [...], а не пределните глаголи са само несвършени“ (Ницолова 2008: 248).

II.2.1. Морфологичните категории лице, число и род при глаголите в български

Морфологичните категории, които притежават глаголите в български език и са обект на анализ в това изследване, са: лице, число, (в някои случаи) род, време, вид на действието и таксис. Акцентът е върху категориите, които изграждат хиперкатегорията време, с оглед на преноса на граматикализираната информация при превод. Накратко ще бъдат представени категориите лице, число и род.

Категорията лице при глаголите характеризира предмета (явлението), „на който се приписва динамичният процесуален признак, от гледна точка на отношението му към речевия акт“ (Ницолова 2008: 225). В българското езикознание се е наложило схващането, че категорията лице е тричленна и е изградена върху опозицията комуникатор – некомуникатор (Герджиков 1984: 222). Признакът комуникатор обединява първо и второ лице, а трето лице притежава признака некомуникатор.

Категорията число при глаголите изразява „характеристика на обективните количествени признаци на явленията“ (предметите), „на които се приписва динамичният признак, то е количествена характеристика на глаголното лице“ (Ницолова 2008: 225). При синтетичните и аналитичните форми на личните глаголи, изразяващи граматичната категория време, има ясни морфологични показатели за различаване на единствено и множествено число.

Част от аналитичните глаголни форми за единствено число, които съдържат в състава си минало свършено деятелно причастие (или минало страдателно причастие), са натоварени с допълнително граматично значение – род.

II.2.2. Особености на българската темпорална система

„Зараждането“ на описанието на съвременната българска темпорална система може да се отнесе към докторската дисертация на Л. Андрейчин (Андрейчин 1938) и по-късната му „Основна българска граматика“ (Андрейчин 1944). В тези свои трудове Л. Андрейчин ясно описва значението резултативност при перфектните времена и значението относителност, както и отношенията между тях и морфологичната категория време, като разделя времената на абсолютни и относителни и на времена с просто и сложно хронологическо отношение (Андрейчин 1944: 208). В по-късни изследвания характеристиките на така описаната темпорална система са подложени на повторно дефиниране и осмисляне.

Авторите на различните концепции от по-ново време са единодушни относно същинското значение на морфологичната категория време и разграничаването ѝ от други значения, носени от българските глаголни форми. Същинското значение на морфологичната категория време може да бъде конструирано само по отношение на определен ориентационен момент (момента на речевия акт или друг момент), който е формалният „темпорален център“, спрямо който отношенията могат да бъдат в привативната опозиция разновременност – неразновременност (Куцаров 2007: 244). В рамките на разновременността допълнително се изгражда опозицията следходност и предходност (Куцаров 2007: 244). Морфологичната категория време в български съдържа следните грамемите (Куцаров 2007: 244):

- грамемата **сегашно време** изразява едновременност на динамичния признак с изказването или липса на информация за разновременност;
- грамемата **минало време** изразява предходност на динамичния признак по отношение на изказването;
- грамемата **бъдеще време** изразява следходност на динамичния признак по отношение на изказването.

Според представената от Ив. Куцаров система (Куцаров 2007: 244) в български език информация за едновременност (или липса на информация за разновременност) на действието и изказването носят формите на сегашно време, минало несвършено време (заедно с информация за относителност), минало неопределено време (заедно с информация за резултативност) и минало предварително време (заедно с информация за относителност и резултативност). Информация за следходност на действието спрямо изказването носят формите на бъдеще време, бъдеще предварително време (заедно с информация за резултативност), бъдеще време в миналото (заедно с информация за относителност) и бъдеще предварително време в миналото (заедно с информация за относителност и резултативност). Информация за предходност на действието спрямо изказването носят формите на минало свършено време.

Семантиката на морфологичната категория вид на действието се изгражда върху „отношението между действие и резултат от действие към един и същ ориентационен момент“ (Куцаров 2007: 249). Съществуват различни виждания за вида на опозицията: еквиполентна (резултативност – акционност) (Герджиков 1984: 115) или привативна

(результативност – нерезультативност) (Куцаров 2007: 250). Информация за вида на действието в български език могат да изразяват следните грамемни (Куцаров 2007: 250):

- грамемата **результативност** изразява резултат от действие, актуализиране на действието в най-близкия следващ ориентационен момент;
- грамемата **нерезультативност** изразява липса на данни за резултат от действие, а в главното си значение сигнализира за действие (акционност).

Според представената от Ив. Куцаров система (Куцаров 2007: 250 – 253) в български резултативност се изразява от формите на минало неопределено време, бъдеще предварително време, минало предварително време (заедно с информация за относителност към минал ориентационен момент) и бъдеще предварително време в миналото (заедно с информация за следходност на резултата и относителност към минал ориентационен момент). Нерезультативност се изразява от сегашно време, минало свършено време, минало несвършено време, бъдеще време и бъдеще време в миналото.

Третата категория, която влиза в състава на граматичната категория (хиперкатегорията) време в българския език, е категорията **таксис** (относителност) и изразява „отношение на действията или резултатите от действия, ориентирани към миналия ориентационен момент (минало време или друг сигнализатор), към действията, неориентирани към миналия ориентационен момент“ (Куцаров 2007: 255). Ив. Куцаров посочва, че относителните форми са ориентирани към минало свършено време или друг негов сигнализатор, докато неотносителните или не са ориентирани, или са ориентирани към момента на изказването, базирайки се на мнението на Г. Герджиков, че „минало свършено време е единственото минало време в българския език“ (Куцаров 2007: 229).

Докато морфологичната категория време се изгражда върху отношението на действие към момента на говорене, категорията таксис се основава на отношението между действие към друго действие (ориентационен момент). Или с други думи, категорията таксис се основава на привативната опозиция относителност – неотносителност (зависим – независим таксис) (Куцаров 2007: 255). Информация за съотнасяне към определен минал момент или липса на такава могат да носят следните грамемни (Куцаров 2007: 255):

- грамемата **относителност** изразява действия (или резултати от действия), ориентирани към единственото минало време в българския език (минало свършено време) или към друг сигнализатор на миналия ориентационен момент;

- грамемата **неотносителност** изразява действия (или резултати от действия), при които липсва информация за отнасяне към някакъв ориентационен момент или (при изразяване на главното значение) са ориентирани към момента на говоренето (абсолютност).

Според представената от Ив. Куцаров система (Куцаров 2007: 253 – 257) в български език относителност се изразява от формите на минало несвършено време, бъдеще време в миналото, минало предварително време (заедно с информация за резултативност) и бъдеще предварително време в миналото (заедно с информация за следходна резултативност спрямо минало изказване – минал ориентационен момент). Липса на данни за относителност носят формите на сегашно време, минало свършено време, бъдеще време, бъдеще време в миналото и минало неопределено време.

II.2.3. Семантична характеристика на българската темпорална система

За представянето на системните отношения и формалното конструиране на значенията на членовете на различните морфологични категории ще бъде използван семантичният модел на Х. Райхенбах (Райхенбах 1947). Представянето на темпоралните отношения в български чрез модела на Х. Райхенбах не е нов подход. Използвана е системата, предложена от Р. Ницолова, в която се въведени следните означения за елементите момент на говорене, интервал на референтност, интервал на действие и момент, в който е наличен резултат от действие (Ницолова 2008: 264 – 265):

T_0 – моментът на говорене (още акт на комуникацията);

R – интервалът на референтност, като е възможно да има повече от един интервал на референтност (R_1);

E – интервалът на действието;

t_{res} – моментът, в който е наличен резултат от действието.

Релациите между изброените елементи са означени по следния начин:

$X < Y$ – X предхожда Y , X е преди Y ;

$X \in Y$ – X е елемент от Y ;

$X \notin Y$ – X не е елемент от Y ;

$X \supset Y$ – X включва Y ;

$X \supseteq Y$ – X включва или съвпада (едновременно е) с Y .

На основата на горните елементи и отношенията между тях се предлага формално-семантично описание на българската темпорална в Таблица 2.

Сегашно време	$T \in R; E \supseteq R$ При сегашно време в български език интервалът на действието включва в себе си интервала на референтност, който пък включва в себе си момента на говоренето.
Минало свършено време	$R < T; R \supseteq E$ Семантиката на миналото свършено време изразява действие, чийто интервал е едновременно с минал интервал на референтност и е затворен в рамките на този интервал.
Минало несвършено време	$R < T; E \supset R$ Минало несвършено време в български означава действие, чийто интервал включва в себе си миналия интервал на референтност.
Бъдеще време	$T < R; R \supseteq E$ Значението на бъдеще време в български език означава действие, което се извършва в бъдещ интервал на референтност, разположен след момента на говорене.
Минало неопределено време	$T \in R; E < R, T_0; t_{res} \in R$ При минало неопределено време се изразява значение, което представя действие, чийто интервал е в плана на миналото, преходен е спрямо интервала на референтност, който включва в себе си момента на говорене.
Минало предварително време	$R < T; E < R < T_0; t_{res} \in R$ Семантиката на минало предварително време означава действие, което се извършва в плана на миналото преди даден минал интервал на референтността.
Бъдеще предварително време	$T < E < R; t_{res} \in R$ Бъдеще предварително време означава действие, което се е извършило преди бъдещ интервал на референтност, в който ще бъде налице резултатът от действието.
Бъдеще време в миналото	$R < E < T$ Бъдеще време в миналото означава действие, което е предстоящо по отношение на минал интервал на референтност.
Бъдеще предварително време в миналото	$R < E < R_1 < T; t_{res} \in R_1$ При формирането на значението на бъдеще предварително време в миналото се използват два интервала на референтност. Единият от тези интервали R е в плана на миналото и действието е предстоящо по отношение на него. Тъй като бъдеще предварително време в миналото е резултативно време, резултатът от действието е наличен във втори интервал на референтност R_1 .

Таблица 2. Формално-семантично описание на българската темпорална система.

II.3. Характеристика на глаголите в английски

Английският език принадлежи към групата на германските езици и въпреки че част от характеристиките му, които са общи с останалите езици от тази група, са нарушени от по-късни интензивни езикови влияния, особено от нормандски, много от общите черти са останали и в съвременния английски език. За разлика от българския език, при който теоретичното осъзнаване на езиковите категории започва сравнително

късно, английското езикознание наследява вековната традиция на класическите граматики, изградени върху класическите средновековни езици – латински и гръцки. Развитието на езикознанието от края на XIX и през целия XX век е белязано от множество научни направления и школи, които дават траен отпечатък върху формалното и функционално представяне на езика, а английският език обикновено се явява основен обект на разглеждане при представянето на различните научни схващания.

В английски език глаголите също се делят във формални лексико-семантични класове и притежават определени морфологични категории (Ранкова и Иванова 1980: 88). Основните лексикални разреди, в които попадат английските глаголи по отношение на своите функционални особености, са спомагателни и пълнозначни глаголи. Глаголите в английски език се делят на преходни и непреходни според семантиката на изразяваното действие: дали действието засяга пряко обект, който се реализира синтактично като пряко допълнение, или не (Валейка и Буиткийн 2003: 89). Съществено деление при английските глаголи е групирането им в два основни граматични класа, според начините, по които се образуват формите им за минало време и минало причастие (Мърфи 1991: 23). Според това деление глаголите в английски език се делят на правилни и неправилни. Формите за минало време и минало причастие на правилните глаголи се образуват чрез прибавяне на окончание *-(e)d* към инфинитива. Неправилните глаголи образуват тези форми по различни формообразователни типове, които могат да бъдат разделени на отделни подтипове (Мърфи 1991: 26).

Морфологичните категории при глаголите в английски, които ще бъдат разгледани, са категориите лице, число, време и вид на действието. Граматична категория род не е глаголна категория в английски и значение за „родовост“ не се изразява посредством а глаголната форма.

II.3.1. Категориите лице и число на глаголите в английски

По отношение на глаголните категории лице и число английският език е загубил голяма част от окончанията си, които изразяват отношението на глаголното лице към изказването. Тези категории съществуват в английския език дотолкова, доколкото глаголните форми задължително се съчетават с подлог, изразен чрез именна фраза или местоимение. С формален маркер за лице и число във формите на синтетичните времена е натоварено само 3. лице, единствено число в сегашно просто време (Филипова 2002:

27). При формите на аналитичните времена със значение за лице и число са натоварени спомагателните глаголи във формите за 1. и 3. лице, единствено число, сегашно продължително и минало продължително време; за 3. лице, единствено число, сегашно перфектно, сегашно перфектно продължително време и минало перфектно време (Филипова 2002: 67).

В дисертацията е прието виждането, че експлицитно изразеният подлог или личните местоимения не са част от глаголните форми в английски.

II.3.2. Особености на английската темпорална система

Ранното описание на английската темпорална система е съществено повлияно от латинските граматики. Първоначално са дефинирани три времена в английски (Джонсън 1640: 21), като по-късните описания на системата на времената се придържат към това схващане и назовават времената чрез латинските им наименования. За пръв път Р. Уолт споменава за свършени (Perfect) и несвършени (Imperfect) форми (Уолт 1762: 34). Схващането за тричленна или двучленна категория време с 2 или 3 члена на категорията вид на действието се утвърждава през втората половина на XIX и началото на XX век. Виждания за дванадесетчленна темпорална система се появява в началото на XX век с ясно дефиниран инвентар, но с различни наименования. В граматиката на У. Баскервил и Дж. Сеуел за пръв път са изброени 12 темпорални форми – 3 времена с по 4 вида на действието (Баскервил и Сеуел 1895: 152). Съществуват и по-усложнени съвременни системи, в които са описани основни и вторични времена, аспектиални форми и модални форми. За целите на тази работа се възприема схващането за граматичната категория време в английски като хиперкатегория, изградена от морфологичните категории време и вид на действието.

Повечето съвременни модели на глаголните морфологични категории в английски приемат виждането за 3 същински времена – минало, сегашно и бъдеще (Хаус и Стивънс 2000; Луис и Мол 2013). Съществуват също така мнения за 4 същински времена – минало, сегашно, бъдеще и бъдеще в миналото или минало бъдеще време (Ранкова и Иванова 1980; Филипова 2002; Кабакчиев 2003). Диференциалните признаци на морфологичната категория време в английски език са два – предходност и следходност (Кабакчиев 2003: 132). С липсата на информация за предходност или следходност на действието е натоварено сегашно време. С информация за предходност е натоварено минало време, а за следходност – бъдеще време (и бъдеще време в

миналото, което носи информация за следходност на действието спрямо минал ориентационен момент).

В английски език категорията вид на действието е сложна морфологична категория, която съотнася продължителността или завършеността на действието към външен (спрямо самото действие) ориентационен момент (Валейка и Буиткийн 2003: 84). Действието може да бъде продължително, завършено, да липсва информация за това или да бъде продължително, но завършено спрямо всеки един от ориентационните моменти на четирите времена, като тези отношения са независими от лексикалното значение на глагола (Филипова 2002: 42). Някои автори разлагат морфологичната категория вид на действието в английски на две самостоятелни категории – категорията вид (aspect) (Валейка и Буиткийн 2003: 84), която изразява продължителността или липсата на информация за продължителност на действието спрямо определен ориентационен момент, и категорията подредба (order), която отнася завършеността на действието към определен ориентационен момент – подрежда момента на действието преди ориентационния момент (Валейка и Буиткийн 2003: 97). Въпреки това в английското езикознание е прието разглеждането на категорията вид на действието като една категория, изградена от еквиполентната опозиция продължителност – резултативност (Ранкова и Иванова 1980; Филипова 2002; Кабакчиев 2003). Диференциалните признаци на категорията вид на действието в английски език са 2: продължителност и перфективност (завършеност/резултативност) (Кабакчиев 2003: 134). С липсата на информация за продължителност или завършеност са натоварени т.нар. прости времена (Филипова 2002: 69). Продължителните времена са носители на признака продължителност (Филипова 2002: 82), а перфектните времена – на признака завършеност (Филипова 2002: 90). Наличието на двата признака се наблюдава в сложните форми на перфектните продължителни времена (Филипова 2002: 101).

II.3.3. Семантична характеристика на английската темпорална система

За английски отново е приложено описание на темпоралната система чрез системата на Х. Райхенбах. За представянето се използват приетите за български език елементи и отношения, за да бъде възможно формално-семантичното съпоставяне между български и английски.

Формално-семантичното описание на английските времена е представено в Таблица 3.

Сегашно просто време	$E=R=T$ При изграждането на семантиката на това време моментът на говорене, моментът на референтност и моментът на действието съвпадат напълно, без да има информация за завършеност или повтаряемост на действието.
Сегашно продължително време	$T = R; E \supseteq R$ Семантиката на сегашно продължително време изразява действие, чийто момент обхваща момента на референтност, който съвпада с момента на говорене.
Сегашно перфектно време	$E < T = R; t_{res} \in R$ Сегашно перфектно време изразява действие в миналото, което е преди момента на говорене. Сегашно перфектно време е резултативно време, като резултатът е наличен към момента на говоренето, който съвпада с момента на референтност.
Сегашно перфектно продължително време	$T \in R; E < R; t_{res} \in R$ Семантиката на сегашно перфектно продължително време се изгражда върху момент на действие, който е преди момента на референтност и говорене, които съвпадат. Резултатът от действието е наличен в рамките на момента на референтност, без информация за приключване на действието.
Минало просто време	$E = R < T$ Семантиката на минало просто време изразява момента на действието, който съвпада с минал момент на референтност, който е преди момента на говорене.
Минало продължително време	$R < T; E \supset R$ При минало продължително време моментът на референтност е преди момента на говоренето, а моментът на действието преминава през него – началото на действието е преди момента на референтност и липсва информация за прекратяване на действието в рамките му, както и към момента на говорене.
Минало перфектно време	$E < R < T; t_{res} \in R$ Семантиката на минало перфектно време в английски език се изразява чрез момента на действието, който е предхождащ спрямо момента на референтност в миналото, но резултатът от действието е наличен в момента на референтност.
Минало перфектно продължително време	$R \supseteq E; R < T; t_{res} \in R$ Минало перфектно продължително време изразява действие, чийто момент е преди минал интервал на референтност, а резултатът от действието е наличен в интервала на референтност. При глаголните форми на това време липсва информация за прекратяване на интервала на действието в рамките на интервала на референтност или преди него.
Бъдеще просто време	$T < R = E$ Семантиката на бъдеще просто време е изградена от момент на референтност, в който се съдържа интервалът на действието и който е след момента на говорене.
Бъдеще продължително време	$T < R; R \supseteq E$ При бъдеще продължително време моментът на референтност е следхождащ по отношение на момента на говоренето, а моментът на действието е включен в него.
Бъдеще перфектно време	$T < E < R; t_{res} \in R$ При бъдеще перфектно време моментът на действието е след момента на говорене, а резултатът от действието е наличен в бъдещ интервал на референтност, който е след момента на говорене и интервала на действието.

Бъдеще перфектно продължително време	$R \supseteq E ; T < E; t_{res} \in R$ Семантиката на бъдеще перфектно продължително време се изгражда от момент на референтност, който е бъдещ спрямо момента на говоренето, и резултат от действие, който е наличен в момента на референтност. Моментът на действието е бъдещ спрямо момента на говорене.
Бъдеще просто време в миналото	$R < E < T$ При бъдеще просто време в миналото моментът на действието е следходен спрямо един минал момент на референтност.
Бъдеще продължително време в миналото	$T > R > R_1; E \supset R_1$ При бъдеще продължително време в миналото в английския език и двата момента на референтност са преди момента на говоренето, единият момент предхожда другия, а момента на действието е бъдещ спрямо първия интервал на референтност и достига до втория – началото на действието е преди втория интервал на референтност и липсва информация за прекратяване на действието в рамките му.
Бъдеще перфектно време в миналото	$R < E < R_1 < T; t_{res} \in R_1$ Семантиката на бъдеще перфектно време в миналото се изгражда от два последователни момента на референтност, предхождащи момента на говорене. Моментът на действието се намира между двата момента на референтност, като спрямо първия е бъдещ, а резултатът от действието е наличен във втория момент на референтност.
Бъдеще перфектно продължително време в миналото	$R < E < R_1 < T; E \supset R_1 t_{res} \in R_1$ Значението на бъдеще перфектно продължително време в миналото се изгражда от два последователни момента на референтност, минали спрямо момента на говорене. Моментът на действието е бъдещ спрямо първия момент на референтност, като достига и/или обхваща изцяло втория, като резултатът от действието е наличен във втория момент на референтност и в него липсва информация за прекратяване на действието.

Таблица 3. Формално-семантично описание на английската темпорална система.

II.4. Сравнителен преглед на характеристиките на глаголните системи и начините за образуване на темпоралните форми в български и английски

Съответствията между темпоралните категории в български и английски са изведени на базата на основната употреба на конкретното време, която се определя от позициите на момента на говорене (T), момента, в който има наличен резултат от действие (T_{res}), интервалите/моментите на референция (R, R_1, R_2) и интервала/момента на действието (E) върху темпоралната ос. Сравнението имат за цел да представят основната употреба на времената. Също така формалният анализ ще бъде ограничен до представяне на отношения между елементите T, T_{res} , R и E. Изчерпателното описание на всички възможни транспозиционни значения и формалните съвпадения между значенията им е извън целите на настоящото изследване. Отношенията между момента на говорене, момента/интервала на действието и момента/интервала на референтност могат да бъдат представени така:

– отношение на съвпадение между момента на говорене, интервала на действието и интервала на референтност ($E=R=T$) в двата езика се изразява от семантиката на формите за сегашно време в български език и формите за сегашно просто и сегашно продължително време в английския език. Основната разлика е в рамките на интервала на действието спрямо интервала на референтност. В български език при значението на сегашно време интервалът на действието е включен в интервала на референтност, докато при значението на сегашно просто време в английски има пълно съвпадение между интервала на действието и интервала на референтност;

– отношение на съвпадение на интервала на действието с интервала на референтност и отношение на предхождане на интервала на референтност спрямо момента на говорене ($E=R<T$) изразяват формите на минало свършено и минало несвършено време в български език и формите на минало просто и минало продължително време в английски език. При значението на минало несвършено време в български интервалът на действието включва в себе си миналия интервал на референтност, а при минало свършено време – интервалът на действието съвпада с минал интервал на референтност. При значението на минало просто време в английски интервалът на действието съвпада изцяло с минал интервал на референтност, докато при минало продължително време моментът на референтност е преди момента на говоренето, а моментът на действието преминава през него;

– отношение на съвпадение между интервалите на действие и референтност и отношение на следхождане на интервала на референтност спрямо момента на говорене ($E=R>T$) изразяват бъдеще време в български език и бъдеще просто и бъдеще продължително време в английски. При бъдеще продължително време интервалът на действието преминава през интервала на референтност като липсва информация за прекратяване на действието в рамките на интервала на референтност.

Отношенията между момента на говорене, момента/интервала на действието, момента/интервала на референтност и момента, в който е наличен резултат от действието, могат да бъдат обобщени по следния начин:

– отношение на съвпадение между интервала на референтност, момента на говорене и момента на наличие на резултат от действие и отношение на предхождане на интервала на действие ($E < R = T = T_{res}$) в български език изразяват формите на минало неопределено време, а в английски – формите на сегашно перфектно и сегашно перфектно продължително време. При сегашно перфектно продължително време има наличие на информация за предхождане на началото на интервала на действие спрямо интервала на референтност, но липсва информация за неговия край;

– отношение на предходност на интервала на действието спрямо интервала на референтност и отношение на предходност на интервала на референтност спрямо момента на говорене и на съвпадение спрямо момента на наличие на резултат от действие ($E < R = T_{res} < T$) в български език изразява семантиката на минало предварително време, а в английски – семантиката на минало перфектно и минало перфектно продължително време. При минало перфектно продължително време има наличие на информация за предхождане на началото на интервала на действие спрямо интервала на референтност, но липсва информация за край;

– отношение на предходност на интервала на действието спрямо интервала на референтност и отношение на следходност на интервала на референтност спрямо момента на говорене и на съвпадение спрямо момента на наличие на резултат от действие ($E < R = T_{res} > T$) в български език изразява семантиката на бъдеще предварително време, а в английски – семантиката на бъдеще перфектно и бъдеще перфектно продължително време. При бъдеще перфектно продължително време има наличие на информация за предхождане на началото на интервала на действие спрямо интервала на референтност, но липсва информация за край.

Отношенията между представените по-горе елементи – момента на говорене, момента/интервала на действието, момента/интервала на референтност, момента, в който е наличен резултат от действието, и втори интервал на референтност – в двата езика могат да бъдат представени така:

– отношение на следходност на интервала на действие спрямо първи интервал на референтност, който се намира в отношение на предходност спрямо

втори интервал на референтност, който от своя страна е в отношение на предходност спрямо момента на говорене ($R > E > R_1 > T$), изразяват значенията на формите за бъдеще време в миналото в български език и бъдеще просто време в миналото и бъдеще продължително време в миналото в английски език. При бъдеще продължително време в миналото има наличие на информация за предхождане на началото на интервала на действие спрямо втория интервал на референтност, но липсва информация за край;

– отношение на следходност на интервала на действие спрямо първи интервал на референтност, който се намира в отношение на предходност спрямо втори интервал на референтност, който от своя страна е в отношение на предходност спрямо момента на говорене и в отношение на съвпадение с момента на наличие на резултат от действие ($R > E > R_1 = T_{res} > T$), в български език изразяват формите на бъдеще предварително време в миналото, а в английски – формите на бъдеще перфектно време в миналото и бъдеще перфектно продължително време в миналото. При бъдеще перфектно продължително време в миналото има наличие на информация за предхождане на началото на интервала на действие спрямо втория интервал на референтност, но липсва информация за край.

По отношение на граматичните значения на категориите, които изграждат темпоралните отношения в български и английски могат да бъдат направени следните заключения:

- глаголни форми, които изразяват едновременност или липса на информация за разновременност на действието с момента на говорене, ненатоварени с други значения, притежават и двата езика. За български това са глаголните форми за сегашно време, за английски – формите за сегашно просто и сегашно продължително време (носеци информация за продължителност на действието);

- глаголни форми, които изразяват информация за следходност на действието спрямо момента на говорене, ненатоварени с други значения, притежават и двата езика. За български това са формите за бъдеще време, за английски – формите за бъдеще просто и бъдеще продължително време (носеци информация за продължителност на действието);

- глаголни форми, които изразяват следходност на резултат от действие спрямо момента на говорене, притежават и двата езика. За български това са формите на бъдеще предварително време, за английски – формите на бъдеще перфектно и бъдеще перфектно продължително време (носеци информация за продължителност на действието);
- глаголни форми, които изразяват предходност на действие спрямо момента на говорене, притежават и двата езика. В български това са формите на минало свършено време, в английски – формите на минало просто време.
- глаголни форми, които изразяват едновременност на резултат от действие с момента на говорене, притежават и двата езика. В български това са формите на минало неопределено време, в английски – формите на сегашно перфектно и сегашно перфектно продължително време;
- глаголни форми, които изразяват наличие на резултат от действие спрямо минал момент, притежават и двата езика. В български това са формите на минало предварително време, в английски – формите за минало перфектно и минало перфектно продължително време;
- глаголни форми, които изразяват следходност на действие спрямо минал ориентационен момент, притежават и двата езика. В български това са формите за бъдеще време в миналото, в английски – формите за бъдеще просто време в миналото и бъдеще продължително време в миналото;
- глаголни форми, които изразяват следходност на резултат от действие спрямо минал ориентационен момент, притежават и двата езика. В български това са формите за бъдеще предварително време в миналото, в английски – формите за бъдеще перфектно време в миналото и бъдеще перфектно продължително време в миналото.

Формалните съответствия между български и английските глаголни форми, както и броя на възможните форми, изразяващи лице и число, са представени в Таблица 4.

Български език	Английски език	Брой форми
Сегашно време	Сегашно просто/Сегашно продължително	
1. лице, ед. число; 2. лице, ед. число; 3. лице, ед. число; 1. лице, мн. число; 2. лице, мн. число; 3. лице, мн. число;	3. лице, ед. число; 1/2. лице, ед. число/1/2/3 лице, мн.число/ 1. и 3. лице, ед. число; 2 лице, ед. число/1/2/3 лице, мн.число	6:2/6:3
Минало свършено време	Минало просто време	
1. лице, ед. число; 2/3. лице, ед. число; 1. лице, мн. число; 2. лице, мн. число; 3. лице, мн. число;	1 форма за лице и число	5:1
Минало несвършено време	Минало продължително време	
1. лице, ед. число; 2/3. лице, ед. число; 1. лице, мн. число; 2. лице, мн. число; 3. лице, мн. число;	1. и 3. лице, ед. число; 2 лице, ед. число/1/2/3 лице, мн. число	5:2
Минало неопределено време	Сегашно перфектно време/Сегашно перфектно продължително време	
1. лице, ед. число, м.р.; 1. лице, ед. число, ж.р.; 1. лице, ед. число, ср.р.; 2. лице, ед. число, м.р.; 2. лице, ед. число, ж.р.; 2. лице, ед. число, ср.р.; 3. лице, ед. число, м.р.; 3. лице, ед. число, ж.р.; 3. лице, ед. число, ср.р.; 1. лице, мн. число; 2. лице, мн. число; 3. лице, мн. число;	3. лице, ед. число; 1/2 лице, ед. число/1/2/3 лице, мн. число	12:2
Минало предварително време	Минало перфектно/Минало перфектно продължително	
1. лице, ед. число, м.р.; 1. лице, ед. число, ж.р.; 1. лице, ед. число, ср.р.; 2/3. лице, ед. число, м.р.; 2/3. лице, ед. число, ж.р.; 2/3. лице, ед. число, ср.р.; 1. лице, мн. число; 2. лице, мн. число; 3. лице, мн. число;	1 форма за лице и число	9:1
Бъдеще време	Бъдеще просто време/Бъдеще продължително време	
1. лице, ед. число; 2. лице, ед. число; 3. лице, ед. число; 1. лице, мн. число; 2. лице, мн. число; 3. лице, мн. число;	1 форма за лице и число	6:1
Бъдеще предварително време	Бъдеще перфектно време/Бъдеще перфектно продължително време	
1. лице, ед. число, м.р.; 1. лице, ед. число, ж.р.; 1. лице, ед. число, ср.р.; 2. лице, ед. число, м.р.; 2. лице, ед. число, ж.р.; 2. лице, ед. число, ср.р.; 3. лице, ед. число, м.р.; 3. лице, ед. число, ж.р.; 3. лице, ед. число, ср.р.; 1. лице, мн. число; 2. лице, мн. число; 3. лице, мн. число;	1 форма за лице и число	12:1
Бъдеще време в миналото	Бъдеще просто време в миналото	
1. лице, ед. число; 2. лице, ед. число; 3. лице, ед. число; 1. лице, мн. число; 2. лице, мн. число; 3. лице, мн. число;	1 форма за лице и число	6:1
Бъдеще предварително време в миналото	Бъдеще перфектно време в миналото/Бъдеще перфектно продължително време в миналото	
1. лице, ед. число, м.р.; 1. лице, ед. число, ж.р.; 1. лице, ед. число, ср.р.; 2. лице, ед. число, м.р.; 2. лице, ед. число, ж.р.; 2. лице, ед. число, ср.р.; 3. лице, ед. число, м.р.; 3. лице, ед. число, ж.р.; 3. лице, ед. число, ср.р.; 1. лице, мн. число; 2. лице, мн. число; 3. лице, мн. число;	1 форма за лице и число	12:1

Таблица 4. Отношение между формите, изразяващи граматичната категория време в български и английски.

III. Създаване на „Българско-английския паралелен корпус със съотнесени глаголни форми“

В тази глава се предлага кратък преглед на развитието на машинния превод като дисциплина и съвременните теории и практики при създаването на лингвистични ресурси за целите на статистическото езиково моделиране и статистическия машинен превод. Ще бъде представен и основният резултат от изследването – „Българско-английският паралелен корпус със съотнесени глаголни форми“.

III.1. Машинен превод – цели и историческо развитие

Под термина машинен превод може да се разбира практическото приложение на компютрите за превод на текстове от един естествен език на друг (Хътчинс 1995: 431). Подходите и методите, използвани при машинния превод, могат да бъдат разделени на три типа, които следват историческото развитие на научното поле. Първият (и исторически най-старият) тип е т. нар. подход на „директен превод“ (direct translation). При този подход системите за машинен превод са специално създадени за определена двойка езици (Хътчинс 1995: 433). Системите „за директен превод“ обикновено са изградени от ръчно написани правила за анализ, генериране и преобразуване на текстове и достатъчно голям двуезиков речник. Като продължение на системите за директен превод се появява вторият тип системи за машинен превод – системите с „трансферни правила“, при които се осъществява последователно надграждане на различните равнища на изходния език, за да се представи междинно абстрактно работно състояние между изходния и целевия език. С натрупването на изчислителна мощ на компютрите и развитието на съвременните схващания за същността на езиковите проявления се появяват системите за статистическо моделиране на езика. Статистическото езиково моделиране се съсредоточава над крайните продукти на езика, за да изследва естествените процеси в него и проявленията им (Коен 2010: 74). Статистическият машинен превод има за цел да представи вероятността за съответствие на две езикови единици в два или повече различни езика. Статистическите системи за машинен превод могат да бъдат описани като функцията, която максимализира продукта между логическото и семантично съответствие на изречение от изходния и изречение от целевия език (S и T) и граматичността (принадлежността) на T по отношение на целевия език. През XXI век успешно се развива машинният превод, базиран на невронни мрежи. Основната структура на невронните мрежи за машинен превод се състои от два повтарящи се слоя от неврони. Входният слой (анализаторът) анализира входните

текстови единици с променлива дължина и ги кодира във вектор с фиксирана дължина. Втората невронна мрежа (декодерът) генерира целевите текстови единици на базата на вектора, генериран от първата мрежа (Фират и др. 2016: 866).

Както традиционните статистически системи за машинен превод, така и невронните системи разчитат преди всичко на достатъчно обемни и надеждни езикови корпуси, върху които да се обучават и оценяват.

III.2. Езиковите корпуси като ресурс за машинен превод

Според определението, дадено от Европейската асоциация на езиковите ресурси, с термина езиков ресурс се означава системата от езикови данни или описанията на такива данни, които са достъпни за машинна обработка с различна цел. Езиковите ресурси може да се класифицират според структурата, в която са представени езиковите данни, или според конкретното си съдържание. От гледна точка на начина, по който са структурирани, езиковите ресурси могат да представляват списъци от думи, корпуси, електронни речници, тезауруси, тематични речници или бази от данни. Корпусите представляват колекция от текстове в електронна форма, подбрани чрез външни критерии, така че да представят, доколкото е възможно даден език или негова разновидност и да предоставят данни за лингвистични изследвания (Синклер 2005 – цит. по Коева 2014: 29). Определенията за езиков корпус обикновено подчертават критериите при подбора на корпусните единици, значението на корпусите за различни лингвистични изследвания, както и възможността, която корпусите предоставят за компютърна обработка (Коева 2014: 29).

Корпусите се представят чрез традиционно възприетите характеристики за езиков корпус – балансираност, представителност и илюстративност (Лийч 1991: 10). Представителността на езиковия корпус се определя със степента, в която той може да обхваща пълния диапазон от проявления на даден езиков феномен. Балансираността на корпуса зависи от текстовата типология – класификацията и характеристиките на включените текстови типове. Илюстративността (подбора на текстовете) за даден езиков корпус е обвързана с представителността и балансираността му.

III.3. Мотивация за избора на корпусния подход при изследването на българско-английските граматични паралели

Темпоралните системи в български и английски споделят основна характеристика – състоят се от категории, които са структурно организирани и могат да се разглеждат като членове на хиперкатегория. По-големият брой възможни морфологични категории, следователно и по-големият брой на възможните граматикализирани значения в български език допринася за по-голямата многозначност по време на превода поради факта, че граматикализираната информация от изходния език трябва да бъде намалена или неравномерно разпределена между различни граматични категории на целевия език. Независимо от това характеристиките на българските и английските темпорални форми имат много съпоставими характеристики (Лазаров 2016а): и двата езика притежават морфологична категория време, която изразява отношението на действието спрямо момента на говорене; в български и английски има наличие на морфологична категория, която изразява отношението на резултат от действие спрямо ориентационен момент – в български това е категорията вид на действието (резултативност), а в английски това е сложна морфологична категория (също наречена вид на действието), която може продължителност и (или) перфективност; във формообразуването на двата езика се използват спомагателни глаголи, които носят част от граматичното значение на глаголната форма; морфологичните категории, изграждащи темпоралните значения, в двата езика могат да бъдат разглеждани като част от хиперкатегория.

По-големият брой възможни граматични категории, следователно и по-големият брой на възможните граматикализирани значения в български език допринася за по-голямата многозначност по време на превода, поради факта, че граматикализираната информация от изходния език трябва да бъде намалена или неравномерно разпределена между различни морфологични категории на целевия език (или да бъде изразена на лексикално и синтактично ниво). Поради характеристиките на двата езика голяма част от граматикализираната информация се предава чрез различни езикови средства и на различни езикови равнища при семантичния трансфер при превод от български на английски.

Една от основните цели на дисертацията е създаването и описанието на преводен модел, съдържащ информация за преводното съотнасяне на българските глаголни

форми с английските им съответствия. За осъществяването на поставената цел е необходимо: да бъде създаден достатъчно голям паралелен корпус, аотиран с глаголните форми в български и техните преводни съответствия на английски; да бъдат извлечени и анализирани статистически данни от аотирания корпус и да се направят експерименти с преводни модели, които са обогатени с данните от аотирания корпус, за да се провери практическото му приложение. Анализът на честотата на глаголните форми в български и преводните им съответствия на английски предоставя възможност за изследване на спецификите на темпоралните категории и преноса на информацията, която те изразяват при превод.

III.4. „Българско-английски паралелен корпус със съотнесени глаголни форми“

„Българско-английският паралелен корпус със съотнесени глаголни форми“ е създаден през периода 2017 – 2018 година, като целта му е да бъде средство за изследване на граматичните паралели между глаголните системи на български и английски: реализациите на морфологичните категории, влизащи в състава на темпоралните системи на двата езика.

Целите при създаването на Корпуса са основно в две насоки – теоретични и практически. Теоретичните цели са да се подберат и съотнесат обективни количествени и дистрибутивни езикови данни за темпоралните системи на български и английски и изчерпателни (за наличния обем) качествени и количествени езикови проявления за надежден анализ. Практическите цели се свеждат до предоставяне на достоверни лингвистични данни, подпомагащи създаването на различни езикови приложения за компютърна обработка на естествения език като статистически модел за аотиране и статистически модел за превод на глаголните форми.

III.4.1. Подбор на езиковите ресурси, включени в Корпуса

„Българско-английският паралелен корпус със съотнесени глаголни форми“ съдържа извадки от други езикови ресурси, подбрани по предварително разработени критерии.

Основният критерий, по който са избрани езикови ресурси за включване в „Българско-английския паралелен корпус със съотнесени глаголни форми“, е наличието на качествена морфологична аотация, тъй като тя представлява съществена част от обработката на лингвистичните ресурси с цел извличане на статистически данни и

конструиране на езиков модел. Ресурсите, които са изследвани и оценени, са „Българският национален корпус“, „Българско-английският паралелен корпус“ (Коева и др. 2012а). и „Българско-английският паралелен корпус със съотнесени изречения (и прости изречения в състава на сложното)“¹ (Коева и др. 2012б). „Българско-английският паралелен корпус със съотнесени изречения“ се отличава със следните характеристики:

- подробна морфологична едноезикова и многоезикова анотация, направена с помощта на Българската многокомпонентна система за първична обработка и лингвистична анотация на текстове²;
- многоезикова анотация, която включва съотнасяне на изреченията в двата езика (и на простите изречения в състава на сложното), като съотнасянето е проверено или направено от човек, а проверката и корекцията на автоматичното съотнасяне са извършени със специално разработена програма.

Критериите за подбор на езиковите ресурси могат да бъдат разделени по следния начин:

- според характеристиките, които са възприети за езиков корпус: да бъде представителен по отношение на езиковия феномен, който представя (спецификите на глаголната употреба предполагат небалансираността на текстовите типове, които ще бъдат включени в Корпуса);
- според съвременното състояние на ресурсите: друго основно изискване е ресурсът да представлява актуалната употреба на езика;
- според допълнителната лингвистична информация в наличните ресурси: подходящите ресурси трябва да представляват паралелни българско-английски сдвоени текстове на определено ниво – изречения или параграфи, и да имат анотационен слой с морфологична информация – маркери за част на речта и граматични характеристики;
- според екстралингвистичните характеристики на ресурсите: в метаинформацията на ресурсите трябва да бъде посочено кой е езикът на оригиналния и езикът на преводния текст (посоката на превод в ресурса)³;

¹ <https://dcl.bas.bg/bulenac/>

² <https://dcl.bas.bg/webservices/>

³ Поради липсата на надеждна информация този критерий не се прилага системно при създаването на Корпуса.

- според авторските права: естеството на подбора предопределя и статута на авторските права на ресурсите: те трябва да бъдат освободени от авторски права, авторски права да бъдат предоставени за целите на работата или източникът винаги се цитира.

С оглед на определените критерии изборът на ресурси е ограничен до „Българско-английския паралелен корпус със съотнесени изречения“. С цел осигуряване на многообразие на илюстрираните темпорални форми се събират освободени от авторски права двуезикови документи от интернет и се използват примери от „Българска граматика“ (Bulgarian Grammar) (Ницолова 2017).

III.4.2. Структура на „Българско-английския паралелен корпус със съотнесени глаголни форми“

Корпусът включва избрани части от „Българско-английския паралелен корпус със съотнесени изречения“, документи от интернет и примери от „Българска граматика“, които са параграфи или изречения. Метаинформацията на всяка корпусна единица включва данни за източника (име на файла или URL), изходния и целевия език, датата на събиране и датата на вписване в Корпуса (ноември 2018 – януари 2019), наличието на маркери за част на речта и граматични характеристики и информация за тематичната област (художествени текстове, публицистични текстове, научни текстове, субтитри и примери от „Българска граматика“).

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“. За да бъдат упоменати авторските права на примерите от „Българска граматика“, информацията за източника е включена в метаданните на съответните файлове.

При работата върху дизайна на Корпуса е възприет един от основните методи за постигане на балансираност при изграждането на езиков ресурс – базирането му на вече възприета структура и извличането на случайни извадки от по-голям корпус. По външни критерии са подбрани ограничен брой текстове от интернет, както и специално конструирани за целта изречения с цел изчерпателност на данните в Корпуса. След всяко ново добавяне на езиков материал в Корпуса се извлича съотношението между общия брой на българските глаголните форми и илюстрираните темпорални значения с цел да

се избегне нарастването на примери, включващи еднотипни темпорални форми. По тази причина таксономията на текстовите типове в Корпуса няма единна структура, а е основана на периодична проверка за езиково насищане на Корпуса – при внасянето на поредния обработен текст към него се проверява дали броят на изследваните езикови единици (глаголните форми) ще се промени значително, или ще остане сравнително постоянен.

Ниската фреквентност на някои темпорални форми в български предполага, че при фазите на добавяне на езиков материал към Корпуса и достигането на езиково насищане определените предполагаеми стойности за честотата на тези темпорални форми няма да се променят значително и ще се запазят ниски през целия процес на работата. С други думи – при включването на специално подбран по определени критерии текстов тип, който внася нови срещания на темпорални форми към масива, не може да се очаква, че броят на новите срещания ще продължи да се изменя пропорционално при следващите етапи на работа. В този смисъл „Българско-английският паралелен корпус със съотнесени глаголни форми“ представлява небалансиран езиков корпус.

III.4.3. Анотационни конвенции при създаването на „Българско-английския паралелен корпус със съотнесени глаголни форми“

Лингвистичните данни в Корпуса имат два слоя анотация. Първият слой е слойът на морфологичните характеристики на думите и се състои от маркери, които показват стойностите им за конкретната глаголна форма. За анотация на текстовете от интернет и примерите от „Българска граматика“, както и за уеднаквяване в маркерите в „Българско-английския паралелен корпус със съотнесени изречения“ е използвано софтуерното приложение ТриТагер (TreeTagger). Вторият анотационен слой съдържа информация за аналитичните глаголни форми и съотнасянето на темпоралните категории в двата езика. За добавянето на втория слой на анотация е използван свободно достъпният инструмент за работа с езикови ресурси УебАно (WebAnno).

Инвентарът на втория анотационен слой се състои от 26 стойности, които представят информация за граматичната категория време. При добавянето на маркерите от този слой са възприети следните принципи:

- лингвистичните маркери за български и английски представят членове на хиперкатегорията време в двата езика – 9 за български и 17 за английски (16 за темпоралните форми и 1 за инфинитива);
- лингвистичните маркери от втория анотационен слой не съдържат информация за лице, число или род на глаголните форми, тъй като тази информация се формира от стойностите на маркерите за морфологичните характеристики на думите;
- допустимо е съотнесените форми да не бъдат поравно разпределени в двата езика – т.е. възможно е глаголна форма в български да няма съответния еквивалент в английски (преводът може да бъде осъществен с други части на речта) и обратно;
- маркерите отразяват само темпоралните характеристики на формите, без да отразяват другите граматични характеристики като наклонение, залог, евиденциалност, положителна или отрицателна форма. Отрицателни форми или форми, които не са в активен залог или индикатив, също са маркирани, но са отразени само темпоралните им характеристики.

Вторият анотационен слой е добавен ръчно и/или полуавтоматично чрез софтуерното приложение УебАно, което позволява цялостно използване на функция за автоматично приписване на лингвистична информация на базата на вероятностен модел и последваща корекция.

Корпусът е във CoNLL⁴ формат. Изреченията се състоят от един или повече редове с думи, а редовете с думи съдържат следните колони: пореден номер на токъна; токън; маркер за морфологичната характеристика на думата от първия анотационен слой; маркер за аналитичната глаголна форма, ако токънът е част от такава; идентификационен номер на глаголната форма. Идентификационният номер може да бъде в три различни отношения между двата езика: идентификационният номер на глаголната форма в български да отговаря на идентификационния номер на глаголната форма в английски; идентификационният номер на глаголната форма в български да няма съответния еквивалент в английски, поради превод в английски с други лексикални средства, които не са глаголна форма; идентификационният номер на глаголната форма

⁴ <https://universaldependencies.org/format.html>

в английски да няма съответния еквивалент в български, поради превод на части от българското изречение, които не са глаголни форми, с глаголни форми в английски).

III.5. Характеристика на „Българско-английския паралелен корпус със съотнесени глаголни форми“

В настоящия си вид „Българско-английският паралелен корпус със съотнесени глаголни форми“ представлява неголям, небалансиран, илюстративен (за поставената си цел), двуезиков паралелен корпус с общ обем от 5 070 анотирани и съотнесени глаголни форми и техните преводни съответствия. Българските текстове съставляват 45 843 думи (48,12% от общия обем), а английските – 49 432 думи (51,88% от общия обем) (Лазаров 2019б).

Схематично представени, текстовете в „Българско-английския паралелен корпус със съотнесени глаголни форми“ изглеждат така (Лазаров 2019б: 111).

- По област:
 - 65,5% художествени текстове – разкази, романи, приказки и др., от които 60% представляват текстове от „Българско-английския паралелен корпус със съотнесени изречения“, а 40% – текстове от интернет;
 - 18% публицистични текстове – новини, изявления, съобщения, от които 100% представляват текстове от „Българско-английския паралелен корпус със съотнесени изречения“;
 - 3% научни текстове – изследвания, статии, от които 100% представляват текстове от „Българско-английския паралелен корпус със съотнесени изречения“;
 - 13% филмови субтитри, от които 100% представляват текстове от „Българско-английския паралелен корпус със съотнесени изречения“;
 - 0,5% други текстове, които са примери от „Българска граматика“.
- По посока на превода:
 - 40,1% от текстовете представляват български оригинал с английски превод;
 - 35,4% от текстовете представляват английски оригинал с български превод;

- 24,5% от текстовете представляват преводи на български и английски от оригинал на трети език.
- По произход:
 - 84% от текстовете представляват текстове от „Българско-английския паралелен корпус със съотнесени изречения“;
 - 15,5% от текстовете са текстове, извлечени от интернет;
 - 0,5% от текстовете са конструирани изречения, извлечени от примери от „Българска граматика“.

Корпусът е създаден като част от проекта „Българо-английски граматични паралели с оглед на машинния превод. Обогаляване на статистически модел за превод от български на английски с лингвистична информация“ по Програма за подпомагане на младите учени в БАН.

IV. Анализ на граматичните паралели в „Българско-английския паралелен корпус със съотнесени глаголни форми“

Основната информация, която може да се извлече от Корпуса, е за употребата на темпоралните форми в двата езика поотделно и в съпоставителен план. В тази глава на дисертацията е представен статистически обусловен анализ на данните от „Българско-английския паралелен корпус със съотнесени глаголни форми“ за разпространението на темпоралните форми в български и в английски. Анализирани са английските преводни съответствия на българските темпорални форми, засвидетелствани в Корпуса.

IV.1. „Българско-английският паралелен корпус със съотнесени глаголни форми“ в контекста на изследването

Преди да бъдат представени статистическите данни, извлечени от „Българско-английския паралелен корпус със съотнесени глаголни форми“, ще бъдат представени основните му характеристики. Беше създаден по-малък експериментален корпус, върху който бяха направени опити, свързани с бъдещата работа по Корпуса (Лазаров 2017). Това предварително изследване показва, че глаголните форми в български и английски имат неравномерно разпределение в различните типове текстове и не може да се очаква да демонстрират всички възможни проявления в различен контекст. От друга страна, беше потвърдено, че корпусно базираният подход е подходящ за извличане на статистическа информация за преводните реализации на глаголните форми от български на английски, както и че наличието на определени граматични категории в български като изходен език предопределя наличието на определени граматични категории в английски като целеви език.

Основната информация, която може да се извлече от Корпуса, е за употребата на темпоралните форми в двата езика поотделно и в съпоставителен план. В първия анотационен слой са анотирани само синтетичните глаголни форми, включително изразяващи морфологичната категория време (сегашно време, минало свършено време, минало несвършено време за български и сегашно просто време и минало просто време за английски език), наклонение (изявително за двата езика), залог (активен залог за двата езика). Граматичните характеристики, които вторият анотационен слой приписва, се отнасят до аналитичните темпорални форми в български и английски и не съдържат информация за други морфологични категории. Особеност на втория анотационен слой

е, че липсва съотнасяне между глаголните форми в български език и други части на речта, с които може да се осъществява преводът на английски език (както и обратно).

IV.2. Статистически обусловен анализ на данните за българските темпорални форми в Корпуса

Данните за честотата на срещане на темпоралните форми в български показват следното разпределение в Корпуса: форми на сегашно време се срещат 2 437 пъти; форми на минало свършено време – 1 311 пъти; форми на минало несвършено време – 710 пъти; форми на минало неопределено време – 269 пъти; форми на бъдеще време – 189 пъти; форми на минало предварително време – 164 пъти; форми на бъдеще време в миналото – 47 пъти; форми на бъдеще предварително време в български – 10 пъти; форми на бъдеще предварително време в миналото – 5 пъти.

По отношение на темпоралните морфологични категории се наблюдава следното разпределение: форми, маркирани със значение неразновременност, се срещат 3 580 пъти; форми, маркирани със значение предходност, – 1 311 пъти; форми, маркирани със значение следходност, – 251 пъти; форми на нерезултативни времена се срещат 4 694 пъти; форми на резултативни времена – 448 пъти; форми на неотносителни времена се срещат 4 216 пъти; форми на относителни времена – 926 пъти.

Представените данни показват, че формите за сегашно време съставляват почти половината от аотираните темпорални форми за български език в Корпуса. Най-рядко срещани са формите на бъдеще предварително време в миналото, които съставляват 0,1% от всички темпорални форми в Корпуса. Прави впечатление, че приоритетно разпределение имат немаркираните членове на съответните темпорални морфологични категории: неразновременност, нерезултативност и неотносителност. Тази особеност може да се обясни с факта, че немаркираните членове могат да изразяват както главното си значение – противоположно на значението на маркирания член, така и общото значение – когато няма информация за наличие или отсъствие на основното значение на маркирания член. Също така се вижда, че 86,7% от българските темпорални форми в Корпуса са синтетични, а останалите 13,3% са аналитични.

IV.3. Статистически обусловен анализ на данните за английските темпорални форми в Корпуса

Честотното разпределение в Корпуса на темпоралните форми при превода на английски е следното: форми на минало просто време се срещат 2 008 пъти; форми на

сегашно просто време – 1 430 пъти; форми на инфинитив – 294 пъти; форми на бъдеще просто време – 264 пъти; форми на минало перфектно време – 159 пъти; форми на минало продължително време – 117 пъти; форми на бъдеще просто време в миналото – 103 пъти; форми на сегашно перфектно време – 96 пъти; форми на сегашно продължително време – 73 пъти; форми на минало перфектно продължително време – 15 пъти; форми на бъдеще перфектно време в миналото – 12 пъти; форми на бъдеще перфектно време – 11 пъти; форми на сегашно перфектно продължително време – 7 пъти; форми на бъдеще продължително време в миналото – 3 пъти; форми на бъдеще продължително време – 2 пъти; не се срещат форми на бъдеще перфектно продължително време и форми на бъдеще перфектно продължително време в миналото.

Разпределението на темпоралните морфологични категории в Корпуса в английски е следното: форми на прости времена се срещат 3 805 пъти; форми на перфектни времена – 278 пъти; форми на продължителни времена – 195 пъти; форми на перфектни продължителни времена – 22 пъти; инфинитивни форми – 294 пъти.

Данните показват, че формите на минало просто време са най-разпространени в английската част на Корпуса, а формите на бъдеще перфектно продължително време и бъдеще перфектно продължително време в миналото не са реализирани в рамките на Корпуса. В Корпуса приоритетно се срещат формите на простите (неперфектни, непродължителни) времена, докато най-комплексните форми на перфектните продължителни времена имат най-ниска честота на разпространение. В английски език на второ място по разпространение са инфинитивните форми на глаголите. От семантична гледна точка инфинитивните форми изразяват действието само по себе си, без връзка с глаголното формоизменение (Ранкова и Иванова 1980: 154), от което следва, че необходимата информация за пълноценен превод не се предава на морфологично равнище.

IV.4. Статистически обусловен анализ на превода на темпоралните форми от български на английски

Преводните реализации в английски на формите на българските темпорални категории, които се срещат в Корпуса, са представени обобщено в Таблица 5.

Времена в български Времена в английски	Сегашно време	Минало свършено време	Минало несвършено време	Минало неопределено време	Бъдеще време	Минало предварително време	Бъдеще време в миналото	Бъдеще предварително време	Бъдеще предварително време в миналото
Сегашно просто време	1346	35	14	14	17	3	2	–	–
Сегашно продължително време	59	3	4	1	6	–	–	–	–
Сегашно перфектно време	29	21	3	42	–	1	–	–	–
Сегашно перфектно продължително време	4	1	–	2	–	–	–	–	–
Минало просто време	289	1084	444	134	6	44	3	–	–
Минало продължително време	21	15	64	8	5	4	1	–	–
Минало перфектно време	9	21	10	29	–	90	–	–	–
Минало перфектно продължително време	3	1	7	1	–	3	–	–	–
Бъдеще просто време	161	1	1	–	102	–	–	–	–
Бъдеще продължително време	–	–	–	–	2	–	–	–	–
Бъдеще перфектно време	–	–	–	–	–	–	–	10	–
Бъдеще просто време в миналото	23	6	8	11	23	3	29	–	–
Бъдеще продължително време в миналото	1	–	–	–	–	–	4	–	–
Бъдеще перфектно време в миналото	–	–	–	1	–	3	3	–	5
Други части на речта	274	90	120	17	15	13	3	–	–
Инфинитив	210	30	34	8	12	–	–	–	–

Таблица 5. Преводни реализации на формите на българските времена в английски.

Данните в Таблица 5. показват, че формите за сегашно време в български език имат най-голямо разнообразие от преводни реализации в английски – 11 темпорални глаголни форми, инфинитивни форми и превод с други части на речта. Значително големият брой от възможни преводни реализации на формите за сегашно време в английски може да бъде обяснен с факта, че в български сегашно време притежава най-голям брой възможни транспозиционни употреби (Ницолова 2008: 274 – 276), някои от които е невъзможно да бъдат преведени с форми за сегашните времена в английски.

Въпреки разнообразието от преводни реализации, доминираща честота имат формите на сегашно просто време в английски (сегашно време в български и сегашно просто време в английски бяха разгледани като семантично най-близки в двата езика на основата на формално-семантичния анализ). Най-ограничени по отношение на броя на преводни реализации са формите на бъдеще предварително време и бъдеще предварително време в миналото – само по една съответстваща темпорална форма. В

рамките на Корпуса не се наблюдават преводни реализации на сегашно време в български език с форми за бъдеще продължително време, бъдеще перфектно време, бъдеще перфектно продължително време, бъдеще перфектно време в миналото и бъдеще перфектно продължително време в миналото в английски език.

Глаголните форми в минало свършено време от българските текстове в Корпуса са 1 311 или 25,50 % от състава на всички български форми. Броят на формите за минало свършено време съставлява около 1/4 от корпуса и 1/2 от броя на формите за сегашно време.

Преводните реализации на формите на минало свършено време в английски също са разнообразни. С най-висока честота са формите за минало просто време в английски, като след тях се нареждат преводните реализации с други части на речта. Формите за минало свършено време имат приблизително същото разнообразие от преводни реализации в английски както формите за сегашно време. Обяснение за разнообразието може да се търси както в транспозиционните употреби, така и при следването на принципа за съгласуване на времената в английски.

В рамките на Корпуса не се срещат преводни английски реализации на формите на минало свършено време с форми на бъдеще продължително време, бъдеще перфектно време, бъдеще перфектно продължително време, бъдеще продължително време в миналото, бъдеще перфектно време в миналото и бъдеще перфектно продължително време в миналото. Семантиката на изброените глаголни форми се обединява около значението следходност на резултат от действие или действие спрямо определен ориентационен момент.

Глаголните форми в минало несвършено време в българските текстове в Корпуса са 710 или 13,81% от състава на всички глаголни форми. Формите за минало несвършено време съставляват приблизително 1/3 от броя на формите за сегашно време и приблизително 1/2 от броя на формите за минало свършено време. При формите на минало несвършено време най-разпространените реализации в английски език – формите на минало просто време, съставляват над половината от общия брой на срещанията.

От представените данни се вижда, че не може да се намери еднозначно съответствие между употребата на минало несвършено време в български и минало продължително време в английски. Най-висока честота при превода на форми на минало

несвършено време в български имат формите за минало просто време в английски, докато реализациите с форми за минало продължително време се нареждат на трето място след превода с други части на речта.

В рамките на Корпуса не се срещат преводни реализации на формите за минало несвършено време в български с форми за бъдеще продължително време, бъдеще перфектно време, сегашно перфектно продължително време, бъдеще перфектно продължително време, бъдеще продължително време в миналото, бъдеще перфектно време в миналото и бъдеще перфектно продължително време в миналото в английски език. Семантиката на изброените по-горе глаголни форми заедно с представените статистически данни дават основание да се направи заключението, че преводни реализации с изброените по-горе глаголни форми в английски на формите за минало несвършено време в български не са вероятни.

Глаголните форми в минало неопределено време от българските текстове в Корпуса са 269 или 5,23% от състава на всички български форми. Формите за минало неопределено време съставляват приблизително 1/3 от броя на формите за минало несвършено време. От представените данни се вижда, че формите за минало неопределено време са с по-ниска честота в български език спрямо формите за сегашно време, минало свършено време и минало несвършено време. Статистическите данни, извлечени от Корпуса, показват нееднозначно поведение при превод на минало неопределено време на английски език – с най-голямо разпределение са преводните реализации с минало просто време в английски език. На второ място по честота са формите за сегашно перфектно време, които са представени в теоретично изведената хипотеза като формален семантичен корелат на формите за минало неопределено време в английски.

Формите на времената в английски, които не се срещат в Корпуса като преводни съответствия на формите на минало неопределено в български, са бъдеще просто време, бъдеще продължително време, бъдеще перфектно време, бъдеще перфектно продължително време, бъдеще продължително време в миналото и бъдеще перфектно продължително време в миналото. Представените данни показват, че не е вероятно форми на бъдещи времена в английски да се използват като преводни съответствия на форми на минало неопределено време в български. Изключение правят бъдеще просто време в миналото и бъдеще перфектно време в миналото поради факта, че те се

използват в подчинени условни изречения, където сигнализируют за възможно приключило действие.

Глаголните форми в бъдеще време от българските текстове в Корпуса са 189 или 3,68% от състава на всички български форми. Формите за бъдеще време съставляват приблизително 1/2 от срещанията на формите за минало неопределено време. От представените данни за разпределението на преводните реализации на формите на бъдеще време се вижда, че приоритетно разпределение имат формите за бъдеще просто време в английски език, което беше представено като формалния семантичен корелат на българското бъдеще време. Разнообразието от преводни реализации е сравнително по-ограничено от реализациите на глаголните форми за сегашно време, минало свършено време, минало несвършено време и минало неопределено време.

Не се срещат преводни реализации на форми за бъдеще време в български с форми за сегашно перфектно време, минало перфектно време, бъдеще перфектно време, сегашно перфектно продължително време, минало перфектно продължително време, бъдеще перфектно продължително време, бъдеще продължително време в миналото, бъдеще перфектно време в миналото и бъдеще перфектно продължително време в миналото на английски. Изброените глаголни форми включват всички перфектни и перфектни продължителни времена, както и бъдеще продължително време в миналото. Семантиката на бъдеще време в български изразява следходност на действие или резултат от действие спрямо момента на говорене. Семантиката на изброените по-горе форми на времена, които не се срещат като преводни еквиваленти в английски на формите за бъдеще време в български, изразява наличие на резултат от приключило или протичащо действие към определен ориентационен момент.

Глаголните форми в минало предварително време от българските текстове в Корпуса са 164 или 3,19% от състава на всички български форми. Формите за минало предварително време почти съвпадат по брой с формите за бъдеще време. От статистическите данни за разпределението на преводните реализации на формите за минало предварително време в български се вижда, че с най-висока честота са формите на минало перфектно време в английски. Формалният семантичен корелат на формите за минало предварително време в български език според представеното формално-семантично описание е минало перфектно време в английски език. На второ място по честота се нареждат формите за минало просто време.

Не се срещат преводни реализации на форми за минало предварително време със следните времена в английски: бъдеще просто време, сегашно продължително време, бъдеще продължително време, бъдеще перфектно време, сегашно перфектно продължително време, бъдеще перфектно продължително време, бъдеще продължително време в миналото и бъдеще перфектно продължително време в миналото. Изброените времена включват формите на всички бъдещи времена в английски и формите на две бъдещи времена в миналото. Предвид семантиката на минало предварително време и изброените по-горе времена, може да се твърди, че преводни реализации с форми за бъдещи времена в английски на форми за минало предварително време в български не са възможни.

Глаголните форми в бъдеще време в миналото от българските текстове в Корпуса са 47 или 0,91% от състава на всички български форми. Броят на формите за бъдеще време в миналото представлява 1/3 от срещанията на формите за минало предварително време. От статистическите данни се вижда, че с най-голяма честота са преводните реализации с форми за бъдеще просто време в миналото в английски език. Формите на бъдеще просто време в български са сред формите на времената с много ниска честота на срещане в „Българско-английския паралелен корпус със съотнесени глаголни форми“ – под 1%.

От данните за разпределението на преводните реализации на формите на времената с ниска честота на срещане в Корпуса може да се направи заключение, че те имат твърде специфична семантика, която се предава чрез ограничен брой форми при превод. Примерите показват, че на значението на формите за бъдеще време в миналото в български отговарят формите за бъдеще просто време в миналото в английски.

Не се срещат преводни реализации на английски на формите на бъдеще време в миналото в български с форми на следните времена: бъдеще просто време, сегашно продължително време, бъдеще продължително време, сегашно перфектно време, минало перфектно време, сегашно перфектно продължително време, минало перфектно продължително време, бъдеще перфектно продължително време и бъдеще перфектно продължително време в миналото. Изброените времена, които не се срещат като преводни съответствия на бъдеще време в миналото, в по-голямата си част представляват минали и сегашни времена.

Форми за бъдеще предварително време и бъдеще предварително време в миналото практически се срещат много рядко в „Българско-английския паралелен корпус със съотнесени глаголни форми“. Всички срещания на форми за тези две времена са единствено в примери от „Българска граматика“. Глаголните форми в бъдеще предварително време от българските текстове в Корпуса са 10 или 0,19% от състава на всички български форми, а формите за бъдеще предварително време в миналото са 5 или 0,1% от състава на всички български форми.

Преводните реализации на формите за бъдеще предварително време и бъдеще предварително време в миналото в примерите от „Българска граматика“ са представени само чрез формите съответно на бъдеще перфектно време и бъдеще перфектно време в миналото. Въпреки това, поради липсата на достатъчно примери не може еднозначно да се определят преводните реализации на тези две времена. Честотата на срещане на формите на бъдеще предварително време в миналото в български представлява доказателство на тезата, че поради сложността на своето значение бъдеще предварително време в миналото се среща в текстовете твърде рядко (Ницолова 2008: 316). Данните от Корпуса могат да служат за същото доказателство и по отношение на бъдеще предварително време.

IV.5. Обобщен анализ на превода на темпоралните морфологични категории от български на английски

В тази част на дисертацията са анализирани статистическите данни, извлечени от „Българско-английския паралелен корпус със съотнесени глаголни форми“, за трансфера на темпоралните морфологични категории, влизащи в състава на хиперкатегорията време в български език, при превод на английски. Представени са обобщено преводните реализации на значенията, изграждащи морфологичните категории време, вид на действието и таксис в български.

Статистическите данни за преводните реализации в английски на формите, изразяващи значенията неразновременност, следходност и предходност в български са представени в Таблица 6.

	Сегашни времена в английски		Минали времена в английски		Бъдещи времена в английски		Инфинитив в английски		Други части на речта в английски	
	Брой форми	%	Брой форми	%	Брой форми	%	Брой форми	%	Брой форми	%
Неразновременност в български език	1532	42,79	1160	32,40	212	5,92	252	7,04	424	11,84
Следходност в български език	28	11,24	12	4,82	179	71,89	12	4,82	18	7,23
Предходност в български език	60	4,59	1121	85,70	7	0,54	30	2,29	90	6,88

Таблица 6. Данни за преводните реализации на английски на значенията, изграждащи морфологичната категория време в български.

Данните показват, че форми, изразяващи неразновременност, се срещат 3 580 пъти в българските текстове в „Българско-английския паралелен корпус със съотнесени глаголни форми“. Реализациите на неразновременност по отношение на морфологичната категория време в английския език са нееднотипни – преобладаващо разпределение имат формите за сегашни времена и формите за минали времена в английски. Глаголни форми, маркирани със значение следходност, се срещат 250 пъти в българските текстове, като при тях преводните реализации са сравнително еднотипни – преимуществено разпределение имат реализациите с форми за бъдещи времена в английски. Глаголни форми, маркирани със значение предходност, се срещат 1 310 пъти, като при тях реализациите също са сравнително еднотипни – с най-висока честота са преводните реализации с форми за минали времена в английски език. Със стабилно поведение при превод се отличават формите, носещи значението разновременност; в английски следходността се реализира най-често с форми за бъдещи времена, а предходността – с форми за минали времена.

Разнообразието на преводните реализации на формите, маркирани със значение неразновременност, може да се обясни с факта, че в български със значение неразновременност са маркирани формите освен на същинското сегашно време, но и на голяма част от „миналите“ времена според класическото схващане за категорията време – минало несвършено време, минало неопределено време и минало предварително време. Семантиката на главните значения на „миналите“ времена може да бъде разделена на два подвида – отнасяща действия или резултат от действия като едновременни спрямо минал ориентационен момент (семантиката на минало несвършено време и минало предварително време) или отнасяща резултати от действия към момента на говорене (семантиката на минало неопределено време). Значението на формите за „миналите“ времена се пренася в английски с форми за минали времена (с

най-висока честота при минало несвършено и минало неопределено време са преводните реализации с минало просто време, а при минало предварително време – с минало перфектно време). От представените статистически данни може да се направи заключението, че макар че са близки по семантика, значенията на морфологичната категория време в български и английски имат различни проявления в темпоралните форми на двата езика.

Статистическите данни за преводните реализации на резултативните и нерезултативните форми от български на английски са представени в Таблица 6.

	Прости времена в английски		Продължителни времена в английски		Перфектни времена в английски		Перфектни продължителни времена в английски		Инфинитив в английски		Други части на речта в английски	
	Брой форми	%	Брой форми	%	Брой форми	%	Брой форми	%	Брой форми	%	Брой форми	%
Резултативност в български	175	37,72	13	2,90	198	49,33	8	1,56	8	1,79	30	6,70
Нерезултативност в български	3630	76,80	182	3,83	80	2,05	14	0,51	287	6,11	502	10,69

Таблица 7. Данни за преводните реализации на английски на резултативните и нерезултативните форми от български.

Данните показват, че нерезултативните форми в български се реализират с най-висока честота в английски чрез форми за прости времена, докато резултативните форми имат нееднотипна преводна реализация, като с най-висока честота са формите на перфектните времена в английски. Глаголните форми, маркирани с признака резултативност в български, могат да бъдат разделени в две подгрупи според разпространението си: времена с относително висока честота на срещане – формите за минало неопределено време и минало предварително време, и времена с много ниска честота на срещане – формите за бъдеще предварително време и бъдеще предварително време в миналото. От представените данни може да се изведе зависимостта, че значението нерезултативност в български език в значителна степен се пренася чрез форми на прости времена в английски, докато значението резултативност има разнородни преводни реализации, като приоритет имат формите за перфектните времена. Тези разнородни реализации се дължат на различната семантика на минало неопределено време (едновременност на резултат с изказването); минало предварително време (едновременност на резултат с минал ориентационен момент); бъдеще предварително време (следходност на резултат към изказването) и бъдеще

предварително време в миналото (следходност на резултат към минал ориентационен момент).

Статистическите данни за преводните реализации на формите на относителните и неотносителни времена са представени в Таблица 8.

	Прости времена в английски		Продължителни времена в английски		Перфектни времена в английски		Перфектни продължителни времена в английски		Инфинитив в английски		Други части на речта в английски	
	Брой форми	%	Брой форми	%	Брой форми	%	Брой форми	%	Брой форми	%	Брой форми	%
Относителни времена в български	544	59,72	73	8,32	116	12,42	7	1,19	34	3,67	136	14,69
Неотносителни времена в български	3261	77,40	122	2,94	162	3,68	15	0,43	260	6,17	396	9,39

Таблица 8. Съотношение на преводните реализации на английски на относителните и неотносителни времена в български.

Статистическите данни показват, че при категорията таксис няма приоритетно разпределение на преводните реализации на английски – относителните и неотносителните времена се реализират чрез форми на прости времена в английски. От данните може да се изведе зависимостта, че категорията таксис в български няма строго определени преводни еквиваленти в английски. Значенията на относителните и неотносителните времена се трансферират при превод чрез значения на различни морфологични категории в английски език. Това може да се обясни с факта, че значението относителност не е разграничено в рамките на отделна морфологична категория в системата на темпоралните отношения в английски.

Статистическите данни за преводните реализации на глаголните форми, образувани с глаголи от свършен и несвършен вид, са представени в Таблица 9.

	Форми на прости времена в английски		Форми на продължителни времена в английски		Форми на перфектни времена в английски		Форми на перфектни продължителни времена в английски	
	Брой форми	%	Брой форми	%	Брой форми	%	Брой форми	%
Глаголни форми с глаголи от свършен вид в български	2674	90,61%	64	2,17%	204	6,91%	9	0,30%
Глаголни форми с глаголи от несвършен вид в български	1131	83,84%	131	9,71%	74	5,49%	13	0,96%

Таблица 9. Съотношение на преводните реализации на английски на глаголите от свършен и несвършен вид в български.

Данните показват, че при глаголите от свършен вид в български приоритетно разпределение с най-висока честота след формите за прости времена имат формите за перфектните времена в английски, а при глаголите от несвършен вид приоритетно разпределение с най-висока честота след простите времена имат формите за продължителните времена в английски. Наблюдаваните зависимости на пръв поглед подкрепят твърдението за еднакво значение на несвършения вид и свършения в български съответно на значението на продължителните и перфектните времена при категорията вид на действието в английски. Честотата на употреба обаче показва, че не може да се търси пряка връзка между лексикално-граматичната категория вид на глагола в български и продължителните и перфектните времена в английски, макар че се наблюдава известно предпочитание за превод на глаголи от несвършен вид в български с форми на продължителни времена в английски, съответно на глаголи от свършен вид в български – с форми на перфектните времена в английски. Независимо от вида на глагола в български, най-голяма честота на разпространение имат формите на простите времена в английски, което доказва тезата за разликата в семантиката на категориите вид на глагола в български и вид на действието в английски.

IV.6. Изводи за българско-английските граматични паралели на темпорални форми

Формулираната хипотеза, че темпоралните морфологични категории в български език системно предопределят морфологичните категории на съответните преводни реализации в английски се потвърждава според представените статистически данни: в 70% от случаите, които са представени, темпоралната форма в български еднозначно предопределя темпоралната форма в английски.

Причината за несъответствието между формално-семантичния анализ и емпиричните данни от Корпуса е комплексна и може да бъде разгледана в две направления – особеностите на двата езика и контекста на употреба. Както беше споменато, в английски език действа правило за съгласуване на времената, което предполага зависима структура на темпоралните форми в изреченията и текста (Ранкова и Иванова 1980: 226). В български език голяма част от времената с по-ниска честота на разпространение изразяват фонови действия, които служат за темпорална ориентация на основното действие в изречението. Съпоставката между разпределението на формите на българските и на английските времена в двете части на Корпуса (българска и английска) също така дава основание за подкрепа на виждането, че честотата на употребата зависи

от семантичната натовареност на членовете на дадена темпорална категория: формите, които изразяват по-комплексни значения, се срещат по-рядко.

В голяма степен темпоралните системи в български и английски и морфологичните категории, които ги изграждат, имат формално-семантични сходства помежду си. Данните от Корпуса показват, че лексикалният контекст, синтактичните особености на езиците, както и прагматичната употреба също влияят при използването на темпоралните форми.

V. Практическо приложение и оценка на „Българско-английския паралелен корпус със съотнесени глаголни форми“

Проведен е експеримент с терминологични бази от данни и софтуер за компютърно подпомогнат превод, които има за цел да провери практическото приложение на лингвистичните данни и статистическата информация от „Българско-английския паралелен корпус със съотнесени глаголни форми“.

Част от „Българско-английския паралелен корпус със съотнесени глаголни форми“ е преработена в терминологична база, за да използва в софтуер за компютърно подпомогнат превод: метоQ. Терминологичната база от данни работи върху предварително аотиран текст с маркери за морфологичните характеристики на думите, който се задава като изходен език за превод в програмата. На базата на маркерите с морфологични характеристики, които присъстват в отделните сегменти (изречения) от изходния език (български), терминологичната база предлага най-вероятния маркер за преводната реализация на глаголната форма от целевия език (английски).

В експеримента участват двама преводачи, които са предварително запознати с тестовата постановка, естеството на текстовете и особеностите на работата със софтуера за компютърно подпомогнат превод. Подборът на текстовете за експеримента е направен така, че да осигури максимално разнообразие от възможни глаголни форми в изходния език. Общият обем на тестовите текстове е 1 580 думи, като в тях се съдържат 200 глаголни форми на български. Текстовете са разделени в две групи – за тестване и за оценка, като за тестване за използвани текстове, съдържащи 150 глаголни форми, а за оценка – текстове, съдържащи 50 глаголни форми. След приписването на най-вероятен маркер на целевия текст на базата на базата данни, извлечена от Корпуса, и входния аотиран текст, е проследено в колко от случаите преводачите ще приемат приписаната автоматично глаголна форма за глагола в текущото изречение или ще я отхвърлят.

В хода на тестването е установено, че в 85% от случаите, в които терминологичната база, изградена върху данните от Корпуса, предлага маркер за глаголната форма в целевия език на базата на езиковите маркери от изходния език, преводачът се съгласява с приписаната глаголна форма. В останалите случаи автоматично предложената глаголна форма не съвпада с избора на преводача, като в 9% от тези случаи избраната от преводача глаголна форма съвпада с втората по вероятност предложена форма, а в останалите 6% – с друга поредна предложена форма. Не е

регистриран случай, в който избраната от преводача глаголна форма да е извън предложения списък с възможни глаголни форми.

Резултатът от проведения експеримент за практическото приложение на Корпуса е, че на базата на моделите за съотнасяне, съдържащи се в Корпуса, могат да се съотнесат автоматично глаголните форми на двата езика, за да се създадат обогатени преводни памети и модели с лингвистична информация, които могат да се използват както за подпомагане на преводния процес, така и за автоматична редакция на вече преведени текстове.

VI. Заключение

Дисертацията предлага теоретичен подход за анализ на свойствата и организацията на темпоралните глаголни системи в български и английски. Избран е изследователски подход, който е доказан като успешен при изследването на съпоставителни проблеми в езици с различна структура. Изследователският подход е базиран на съпоставителното разглеждане на формално-семантичното представяне на темпоралните категории в двата езика и статистически модел за превод на темпорални форми, извлечени от специално разработен за целта паралелен двуезиков корпус: „Българско-английския паралелен корпус със съотнесени глаголни форми“.

VI.1. Обобщение

Представеното изследване е актуално на фона на засиления през последните години интерес към различни научно-приложни разработки за описание и анализ на езиковите явления чрез методите на компютърната лингвистика и в частност чрез възможностите на корпусно базирания подход и статистическото моделиране. Работата е насочена към комплексно и непротиворечиво описание на механизмите за пренос на граматикализираната информация от глаголните форми при превод от български на английски. Изучаването на механизма на взаимодействие между темпоралните системи в български и английски и морфологичните категории, които ги изграждат, има теоретична стойност за изясняване на закономерностите при съотнасянето между граматикализираната информация, семантичната структура и начините за изразяване на темпоралните отношения в двата езика.

Анализът на граматичните паралели в „Българско-английския паралелен корпус със съотнесени глаголни форми“ представя: данните за българските и английските темпорални форми поотделно, данните за трансфера на информация за морфологичните категории време, вид на действието и таксис в български към английските морфологични категории време и вид на действието.

Представените данни свидетелстват, че теоретичната хипотеза за съответствието при формално-семантичния анализ на двата езика е доказуема чрез избрания изследователски подход, и показват, че в съществена част от случаите темпоралната форма в английски като целеви език зависи от комбинаторното наличие на определени морфологични категории в български като изходен език. Значението на контекста за

интерпретацията на глаголните форми в български и английски представлява обективна пречка за пълно потвърждение на хипотезата: при формално-семантичния анализ контекстът не се взема под внимание, докато при анализа на реалните случаи на преводни еквиваленти, регистрирани в „Българско-английския паралелен корпус със съотнесени глаголни форми“, контекстът играе съществена роля за изразяване на темпоралните значения.

От представените статистически данни, разглеждани независимо от теоретичната хипотеза, се вижда, че глаголните форми с най-висока честота на разпространение имат сравнително разнообразни преводни реализации, докато глаголните форми с по-ниска честота на разпространение имат сравнително малък брой възможни преводни реализации в английски. Също така определени темпорални форми не се срещат в рамките на езиковите данни от Корпуса, което доказва тезата за намаляващата им употреба. Могат да бъдат направени следните изводи:

1. Темпоралните системи на българския и английския език имат достатъчно сходства във формално-семантичен план, за да бъдат разглеждани паралелно и да е възможно изграждането на надежден преводен модел, базиран на семантичните свойства на морфологичните категории, които изграждат двете системи.
2. Морфологичните категории, които съставят темпоралните системи в български и английски, имат сходни формално-семантични характеристики и предполагат сходно моделиране на граматикализираните значения в двата езика.
3. Статистическите данни, извлечени от Корпуса, доказват, че комбинаторното наличие на определени морфологични категории в български език, които изграждат темпоралното значение, предопределя темпоралната форма при превод на английски. Различават се начините за изразяване на граматикализираната информация между двата езика. Статистически данни също така показват, че най-често срещаните глаголни форми в български имат сравнително разнообразни преводни реализации в английски, докато по-комплексните и съответно по-рядко разпространени глаголни форми имат сравнително еднотипни преводни реализации.

Изводите спрямо изследователския подход, избран за анализ на обекта на изследването, са:

1. Корпусно базираният подход се явява функционален и успешен за изследването на граматичните паралели между темпоралните системи в български и английски.
2. Обемът и представителността на „Българско-английския паралелен корпус със съотнесени глаголни форми“ са достатъчни, за да дадат отговор на поставените въпроси в това изследване, но не са достатъчни, за да представят пълна картина на различните особености на темпоралните системи в български и английски. Ниската честота на разпространение и твърде специфичните срещания на някои темпорални форми са причина да няма достатъчно надеждни лингвистични данни за тяхното изследване.
3. Провеждането на експеримент с използването на терминологична база от данни, съдържаща преводната граматична информация от Корпуса, за компютърно подпомогнат превод показва успешното приложение на паралелните езикови модели за подобряване на превода. Данните от Корпуса могат да бъдат приложени като помощно средство при работа със системите за компютърно подпомогнат превод, както и при подбора на едоезикови и паралелни данни за обучение на системите за машинен превод.

Изследването представя актуален проблем, свързан с комплексен езиков феномен. Проблемът е представен чрез съвременен и научно доказан изследователски подход, чрез който са реализирани целите на проучването и са предоставени емпирични и количествени данни за обекта на изследване: българско-английските граматични паралели при превода на глаголните форми от български на английски.

VI.2. Приноси на изследването

Научните приноси на изследването може да се обобщят в няколко основни направления, които отговарят на организацията на частите на дисертацията, и могат да бъдат формулирани, както следва:

Представяне на теоретично описание на темпоралните системи в български и английски, както поотделно, така и в съпоставка. Проследено е развитието на теоретичните виждания за категориите, изграждащи темпоралните значения при глаголните форми. В съпоставителен план е изследван механизмът на конструиране на граматичното темпорално значение в българските и английските глаголни форми и са

формулирани основните закономерности при взаимодействието между темпоралните морфологични категории в двата езика. Представено е формално-семантично съпоставяне между значенията на темпоралните категории в български и английски.

Конструиране и аотиране на „Българско-английския паралелен корпус със съотнесени глаголни форми“. При работата по дисертацията не беше намерена информация за наличието на друг такъв паралелен ресурс за български и втори език. Представени са процесът по подбор на ресурсите, предварителната обработка, аотацията и използваните софтуерни приложения за работата по Корпуса.

„Българско-английският паралелен корпус със съотнесени глаголни форми“ предоставя възможности за изследването на граматическите паралели между български и английски език. Корпусът може да се разглежда като 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. Езиковите данни от Корпуса могат да бъдат използвани за обогатяване на преводни модели за нуждата на машинния превод, както и за теоретични изследвания. Представеният работен процес по структурирането и аотирането на Корпуса може да послужи при разработването на подобни лингвистични ресурси за изследване на сходни езикови явления.

Представена е 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 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.

Библиография

- Алексова 2003: Алексова, Кр. Синтагматичната неутрализация на граматическите опозиции като причина за различните състояния, които морфологичната система може да заема (върху материал от темпоралната система на българския език). – В: *Материали от Международния славистичен симпозиум в рамките на Юбилейните Паисиеви четения*. Пловдив: Пловдивски университет [http://belb.info/personal/aleksova/neutralizacii_1.pdf] (посетен на 30.01.2020)
- Андрейчин 1938: Андрейчин, Л. Обща характеристика на българската глаголна система (превод на *Ogólna charakterystyka bułgarskiego system koniugacyjnego*, 1938). – В: Пашов, П. Ницолова, Р. (съст.) *Помагало по българска морфология*. Глагол. София: Наука и изкуство.
- Андрейчин 1944: Андрейчин, Л. *Основна българска граматика*. София: Наука и изкуство.
- Атанасова и др. 1963: Атанасова, Т. Ранкова, М. Русев, Р. Радулова, Н. *Английска граматика в сравнение с български език*. София: Наука и изкуство.
- Баскервил и Сеуел 1895: Baskerville, W. Sewell, J. *An English Grammar: For the Use of High School, Academy, and College Classes*. New York: American Book Company [https://archive.org/stream/anenglishgrammar00sewegoog#page/n156/mode/2up] (посетен на 20.02.2020).
- Валейка и Буиткийн 2003: Valeika, L. Buitkiene, J. *An introductory course in theoretical English grammar*. Vilnius: Vilnius Pedagogical University.
- Вейга 1991: Veiga, A. *Condicionales, concesivas y modo verbal en español*. Santiago de Compostela: Universidade de Santiago de Compostela.
- Гаравалова 2003: Гаравалова, Ил. *Школата на натуралната морфология и българската морфологическа система*. София: СемаРШ
- Герджиков 1984: Герджиков, Г. Преизказването на глаголното действие в българския език. София: Наука и изкуство.
- Герджиков 2000: Герджиков, Г. Категорията време като хиперкатегория – В: сп. *Български език*, бр. 1, 2000 [https://litenet.bg/publish/ggerdzhikov/hyper.htm] (посетен на 20.02.2020).
- Джонсън 1640: Johnson, B. *The English Grammar*. 1640, London [https://archive.org/details/englishgrammar00jons/page/n8] (посетен на 20.02.2020).
- Кабакчиев 2003: Кабакчиев, К. *Английска граматика*. София: Пенсофт.
- Клайн 1994: Klein, W. *Time in Language*. London: Routledge.
- Коева 2014: Коева, Св. Българският национален корпус в контекста на световната теория и практика – В: *Езикови ресурси и технологии за български език*. София: Академично издателство „Проф. Марин Дринов“, 29-53.
- Коен 2010: Koehn, P. *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- Комри 1985: Comrie, B. *Tense*. Cambridge: Cambridge University Press.
- Кударов 2007: Кударов, И. *Теоретична граматика на българския език: морфология*. Пловдив: Университетско издателство „Паисий Хилендарски“.

- Лийч 1991: Leech, G. The state of the art in corpus linguistics. – In: *English Corpus Linguistics*, London: Longman, 8 – 29.
- Луис и Мол 2013: Lewis, G. Mol, H. *Grammar for Young Learners*. Oxford: Oxford University Press .
- Манинг и др. 2013: Manning, L. Cassel, D. Cassel, J. C. St. Augustine’s Reflections on Memory and Time and the Current Concept of Subjective Time in Mental Time Travel. – In: *Behavioral Sciences*, 2013, 3, 232 – 242.
- Мърфи 1991: Murphy, R. *English grammar in use*. Cambridge: Cambridge University Press.
- Ницолова 2008: Ницолова, Р. *Българска граматика: Морфология*. София: Университетско издателство „Св. Климент Охридски“.
- Ницолова 2017: Nicolova, R. *Bulgarian Grammar*. Berlin: Frank & Timme GmbH.
- Паскалева 1987: Паскалева, Е. *Компютър и превод*. София: Наука и изкуство.
- Райхенбах 1947: Reichenbach, H. *Elements of Symbolic Logic*. New York: The Macmillan Company, 1947, 287-298.
- Ранкова и Иванова 1980: Ранкова, М. Иванова, Цв. *Английска граматика*. София: Наука и изкуство.
- Стамболиева 2015: Стамболиева, М. Класове думи: английско-български паралели – В: *Годишник на департамент „Англицистика“*. София: НБУ.
- Уолт 1762: Lowth, R. *A Short Introduction to English Grammar with Critical Notes*. Philadelphia: Aitken
[<https://archive.org/stream/shortintroduction00lowtrich#page/34/mode/2up>] (посетен на 20.02.2020).
- Филипова 2002: Филипова, М. *Практическа английска граматика*. София: ЕМАС.
- Фират и др. 2016: Firat, O. Cho, K. Bengio, Y. Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism. – In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 866–875
- Хаус и Стивънс 2000: House, C.; Stevens, J. *Grammar no problem*, 2000, Berlin: Cornelsen.
- Хорнщайн 1990: Hornstein, N. *As Time Goes By: Tense and Universal Grammar*, Cambridge: MIT Press.
- Хътчинс 1995: Hutchins, W. Machine Translation: A Brief History. – In: Koerner, E. Asher, R. (eds.) *Concise history of the language sciences: from the Sumerians to the cognitivists*. Oxford: Pergamon Press, 431-445.
- Чергова 2012: Чергова, В. *Конюнктивният имперфект в съвременния португалски език*. София: Университетско издателство „Св. Климент Охридски“
- Чергова 2014: Чергова, В. Йерархичност на категориалните значения и форми – В: Кръстев, Н. (съст.) *Мислене, език, реч*. София: Университетско издателство „Свети Климент Охридски“.

Списък с публикациите по темата на дисертацията

- Лазаров 2016а: Лазаров, Т. Особенности на глаголните системи и начините за изразяване на времето в български и английски. Семантичен трансфер при превод на глаголните форми от български на английски – В: „*Littera et Lingua*“, кн. 1, 2016
- Лазаров 2016б: Лазаров, Т. A Possible Solution to the Problem of Machine Translation of Verb Forms from Bulgarian to English – В: *Proceedings of the Second International Conference Computational Linguistics in Bulgaria (CLIB 2016)*, 95 – 100.
- Лазаров 2016в: Лазаров, Т. Анализ на ресурсите за статистически модел за превод на глаголните форми от български на английски – В: *Български език*, 63 (2016), 4. 95 – 101.
- Лазаров 2017: Лазаров, Т. Функционални граматични паралели при превода на глаголните форми от български на английски – В: *Доклади от Международната юбилейна конференция на Института за български език „Проф. Любомир Андрейчин (15/16 май 2017)“*, 321-327.
- Лазаров 2018: Лазаров, Т. Bulgarian-English Parallel Corpus for the Purposes of Creating Statistical Translation Model of the Verb Forms. General Conception, Structure, Resources and Annotation – В: *Proceedings of the Third International Conference Computational Linguistics in Bulgaria (CLIB 2018)*. 193 – 202.
- Лазаров 2019а: Лазаров, Т. Българо-английски паралелен корпус със съотнесени глаголни форми – В: *сп. Съпоставително езикознание, XLIV (2019)*, 1. 5 – 14.
- Лазаров 2019б: Лазаров, Т. Съвременно състояние на Българско-английския паралелен корпус със съотнесени глаголни форми – В: *Български език*, бр. 1, 2019, 103 – 113.