



РЕЦЕНЗИЯ

на диссертационную работу Тодора Димитрова Лазарова,
докторанта по заочной форме к Секции по компьютерной лингвистике, ИБЕ - БАН
на тему „Болгарско-английские грамматические параллели в свете машинного перевода“.

Профессиональное направление 2.1. Филология. Научная специальность: Общее и
сравнительное языковедение (Компьютерная лингвистика).

Научный руководитель: проф. д-р С. Коева

Рецензент: проф. д-р Радка Влахова-Руйкова, СУ им.Св. Климента Охридского

Данные о докторанте

Тодор Лазаров окончил специальность Болгарская филология и Магистерскую программу „Компьютерная лингвистика. Интернет технологии в гуманитаристике“ СУ им.Св. Климента Охридского при Факультете славянских филологий. Он поступил в качестве докторанта очной формы обучения по Общему и сравнительному языковедению (Компьютерная лингвистика), позже форма докторантуры трансформирована в заочную. Докторант обладает отличными дигитальными умениями, владеет 4 языками – английским, русским, немецким, нидерландским.

Во время докторантуры Лазаров принимал участие в Летней школе дигитальной гуманитаристики и прошел две краткосрочные специализации – в Университете Уилверхемптон, Великобритания и в Университете в Осло, Королевстве Норвегия. Во время специализаций он выступал с лекциями по проблематике диссертации. Следует подчеркнуть, что участие докторанта в обучении и специализациях имеет прямое отражение в диссертационной разработке. Впечатление производит и его участие в научных проектах. Он получил приз за лучший проект по Программе о поощрении молодых ученых в 2017 году за проект, которым он руководил.

Т. Лазаров был на стажировке в СКЛ – ИБЕ, БАН, также и лектором в магистерских программах Нового болгарского университета (НБУ), в настоящее время он работает в переводческом агентстве.

Согласно Докладу о выполнении требований Правил для заочных докторантов в БАН докторант выполнил все требования. Следует отметить, что по разделам II. и III. – апробация результатов выполнения научной программы и публикаций по теме, докторант накопил в два раза больше кредитов, чем этого требует минимум. Т. Лазаров выступал с докладами на CLIB 2017, CLIB 2019, на Международной юбилейной конференции Института болгарского языка (ИБЕ), как и на двух международных форумах, организованных Дирекцией „Письменные переводы“ Европейской комиссии (2018, 2019). Публикации – в реферированных и индексируемых изданиях – в журналах „Български език“ („Болгарский язык“), „Сопоставительное языковедение“ („Съпоставително езиковедение“), как и в сборниках материалов международных научных форумов.

Структура текста

Исследование структурировано следующим образом: *Общая характеристика диссертационного труда*, где представлены актуальность и мотивация исследования, цели, объект и предмет и метод исследования; *Болгарско-английские параллели в области грамматической категории времени; Анализ грамматических параллелей в „Болгарско-английском параллельном корпусе со соотнесенными глагольными формами“; Практическое употребление и оценка „Болгарско-английского параллельного корпуса со соотнесенными глагольными формами“; Заключение*, которое включает обобщение и научный вклад, *Библиография и Список публикаций по теме диссертации*. Диссертация содержит 197 с., библиография включает 141 источник болгарских и иностранных ученых, представительные для темы исследования. В Приложении доступен созданный диссертантом корпус.

В введении представлены актуальность проблематики, мотивация и выбор темы, сформулированы основные цели и задачи, указан основной подход. Еще в этой вводящей части Т. Лазаров проявляет себя как исследователя с солидными лингвистическими познаниями и способностью четко формулировать цели и задачи исследования. Тема диссертационна. Исследование имеет теоретический и теоретически-прикладной характер и рассматривает одну из самых актуальных и динамически развивающихся областей компьютерной лингвистики – машинный/ автоматический перевод.

Вторая глава. *Болгарско-английские параллели грамматической категории времени*. В этой главе хочу отметить на первом месте лингвистическую осведомленность докторанта и зрелость выведения основных проблем при сопоставлении двух несимметричных, на первом месте неравноколичественных, систем - 9 граммем в болгарском языке и 16 в английском.

Исследование Т. Лазарова является вкладом в исследования в области болгарского глагола – грамматические категории, которые присущи болгарскому глаголу и грамматические значения, которые он выражает. Подробно и углубленно диссертант представляет различные подходы и концепции болгарских и иностранных языковедов о темпоральной системе глагола в болгарском и в английском языках, как и подходы при наличии асимметрии. Аргументировано и на основе богатого эмпирического материала разграничены формы, выражающие морфологически сигнализированную категорию „вид действия“. На поле болгарского глагола велись настоящие филологические „битвы“, и обычно исследователи глагольных категорий и их значения придерживаются той или другой теории, респективно, описательной модели. Докторанту удалось максимально детально описать морфологически сигнализированные значения граммем, которые объединены в болгарской и в английской темпоральной системе. В работе проводится разграничивание категории *вида действия и темпоральности*. Рассмотрена также и категория *вид действия* в связи с оппозицией *результативность : нерезультативность/аукционность*, как и отношение *таксис : относительность*. Как для болгарского языка, так и для английского в работе сделано целостное описание граммем, объединенных в темпоральную систему. Представлен подробный анализ словоформ, и все они богато иллюстрированы.

В своем исследовании Т. Лазаров не только сопоставляет различные описательные модели, а мотивированно подчеркивает те компоненты значения, которые могут быть формализованными, и их можно включить в сопоставительный анализ с учетом машинного перевода. Разграничены морфологически сигнализированные значения от синтаксических функциональных соответствий, аналитические формы и значения, которые выражаются эквивалентным, лексико-грамматическим путем. В Таблице 4 (с. 78) четко представлены значения граммем в двух языках, а в Таблице 5 (с. 79) соотнесены отдельные словоформы. Производит впечатление, что самая большая разница – 12:1/2, это

разница при формах *прошедшего неопределенного, будущего предварительного и будущего предварительного времени в прошлом*. Разница состоит в основном в наличии только 1 (одной) формы для лица и числа при соответствующих темпоральных формах в английском.

Диссертант рассматривает типологию языков с точки зрения частей речи и места глагола в системе. Очень хорошо проведено разграничение между грамматическим (семантическим по своей природе) и нарративным временем. Время рассмотрено как гиперкатегория, чьи значения реализуются в иерархической системе признаков. Т. Лазаров сопоставляет формообразование в диахронном и в синхронном плане и указывает на основную разницу между двумя языками в отношении категории *вида действия* – морфологической категории для болгарского языка и синтаксической категории для английского. Очень хорошо представлена и роль порядка слов при идентификации подлежащего в английском. Здесь хочу отметить и отношение диссертанта к вопросу о том, следует ли включать личные местоимения в позиции подлежащего в глагольной парадигме как выразители значения *лица* в английском. Независимо от несимметричности форм, которые выражают это значение в болгарском и английском, Т. Лазаров правильно принимает, что нет основания для подобного решения.

Сопоставление значений темпоральных форм в болгарском и в английском, которое базируется на модели Райхенбаха, развитой позже в отношении болгарского глагола Ницоловой, вводящей интервалы, которые, со своей стороны, могут быть открытыми/закрытыми/открытыми только с одной стороны – левой или правой, представлено наглядно в таблице 5., и это сопоставление четко, убедительно и соотносимо с описаниями значений темпоральных форм и у других авторов - Куцаров, Герджиков, Пашов для болгарского языка и с описаниями авторов, исследующих английский глагол. Основные вопросы, сформулированные диссертантом согласно задачам, которые он должен был решить, решены в двух направлениях: относительно трансфера грамматализированной информации из темпоральных глагольных форм – способ передачи этой информации при переводе, и степень трансфера между двумя языками. Основной вопрос в том, можно ли ожидать, чтобы грамматические категории, выражающие сходные значения в двух языках, соответствовали полностью при переводе,

и, соответственно, категории, которые не имеют соответствующих коррелятов, каким образом трансформируют свое значение. Другая группа вопросов – это вопросы относительно типов формообразования и трансфера информации, которую несут соответствующие морфемы – возможно ли и в какой степени наличие определенных морфем, нагруженных соответствующим грамматическим значением в болгарском языке, предполагает наличие соответствующих морфем в английском.

Третья глава *Создание болгарско-английского параллельного корпуса соотнесенных глагольных форм* рассматривает на первом месте историю машинного перевода, снова исчерпывающе и компетентно, и языковые корпуса как ресурсы для машинного перевода. Опираясь на трансферных правилах как на вид грамматики, диссертант обосновывает корпусно базированный подход и статистическое языковое моделирование как самые подходящие для целей исследования (с. 105-106). „Результаты работы по созданию лингвистически обогащенной статистической модели автоматического перевода будут в основном в двух направлениях: теоретическом – выяснение зависимостей при переводе болгарских глагольных форм на английский язык, и практическом – обогащение статистических переводных моделей для автоматического перевода лингвистической информации о переводном соотношении глагольных форм, выражающих различные категории“ (с. 106).

Для выполнения поставленных задач, диссертант создал двухязыковой параллельный корпус соотносящихся темпоральных форм в болгарском и в английском, который иллюстративен (в соответствии с заданной целью). Созданный Т. Лазаровым корпус содержит 5070 аннотированных и соотнесенных глагольных форм и их переводных соответствий. В своем окончательном виде „Болгарско-английский параллельный корпус соотнесенных глагольных форм“ состоит из 84% текстов „Болгарско-английского параллельного корпуса соотнесенных предложений“, 15,5% текстов из интернета и 0,5% предложений, взятых из примеров в „Болгарской грамматике“.

Соотнесенные глагольные формы и их переводные соответствия представлены наглядно и иллюстрированы в таблицах для каждого глагольного времени. Во Второй главе диссертант обращает внимание на наличие конкорданса в английском и отсутствие его в болгарском. Зависимое употребление форм настоящего времени в подчиненном

предложении в болгарском обуславливается событием в главном, которое выполняет роль момента референтности. На этом этапе диссертант не ставит задачу извлечь статистику о том, сколько примеров употреблены в простом и сколько в сложном предложении. Это может быть предметом будущего исследования, так как многие примеры показывают, что информация идентифицируется на этих уровнях. Статистические данные, выведенные диссертантом, являются надежной основой как для машинного перевода, так и для теоретических исследований по глаголу в болгарском и в английском языках. Интересную параллель можно обнаружить и при формах перфекта – 6,05% в болгарском и 6,40% инфинитива в английском.

В ряде наглядных таблиц представлены и анализированы выражаемые темпоральными формами грамматические значения. Они представляют описанные уровни аннотации и соответствующую статистику. Данные о двух языках соотнесены по тем же самым показателям. Диссертант мотивировано выбрал и критерии сопоставления - на базе маркированных и немаркированных форм, респ. с основным значением, на основании описания, воспринятого Куцаровым, и на прагматико-функциональных соответствий, базирующихся на модели Райхенбаха. В этом смысле Т. Лазаров останавливается и на категории вида в связи с длительными временами в английском.

Докторант правильноно заключает, что „Данные Корпуса показывают, что в значительной степени лексикальный контекст и синтаксические особенности языков, как и прагматическое употребление предопределяют использование темпоральных форм. Грамматикализованная информация сохраняется в значительной степени, но переносится также и через дополнительные языковые элементы – через лексикальный материал или контекстную согласованность между отдельными темпоральными формами в рамках текста. Статистические данные так же показывают, что исследование морфологических корреспонденций между темпоральными системами болгарского и английского языков и переводные реализации через „Болгарско-английский параллельный корпус соотнесенных глагольных форм” возможно и приносит пользу.“ (с. 175)

Задача корпуса, по определению диссертанта, - создание обогащенных переводных памятей. Надежным показателем является то, что 85% первого предложения совпадают с

выбором переводчика. Важным результатом также является факт, что нет выбора, который не совпадал бы ни с одним предложением (с. 185). В диссертационном исследовании представлена объективная количественная и дистрибутивная информация о глагольных системах болгарского и английского языков, подходящая для разного типа научного анализа вне охвата исследования. Обеспечены надежные лингвистические данные с целью создания различных языковых приложений для компьютерной обработки естественного языка как статистической модели для аннотации и статистической модели для перевода глагольных форм. Следует подчеркнуть, что лингвистические данные „Болгарско-английского параллельного корпуса соотнесенных глагольных форм “ интегрированы в софтвер, в помощь компьютерного перевода с целью его практического применения в качестве вспомогательного средства при переводе.

Язык текста понятен и точен. Автореферат корректно отражает содержание диссертации, выводы и научный вклад исследования.

В заключение считаю, что диссертант выполнил свои обязанности согласно Правил для докторантов по заочной форме БАН. Представленный к публичной защите текст на тему „Болгарско-английские грамматические параллели в свете машинного перевода“ показывает углубленную лингвистическую подготовку, исследовательский талант, умение создавать научный продукт. Апробация и публикации научных результатов – на авторитетных научных форумах и в реферированных изданиях. Результаты убедительны и хорошо аргументированы. С моей стороны налицо все основания – формально и по существу убежденно поддержать положительным голосованием присуждение образовательной и научной степени ДОКТОР Тодору Димитрову Лазарову по Профессиональному направлению 2.1. Филология, Научная специальность Общее и сопоставительное языковедение (Компьютерная лингвистика).

18. 04. 2020 г.

Подпись:

(проф. Р. Влахова-Руйкова)