**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕСРИТЕТ**

**ФИЛОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ**

образовательная программа магистратуры "Прикладная, экспериментальная и
математическая лингвистика"

РЕЦЕНЗИЯ

На выпускную квалификационную работу студента Горлова Никиты Геннадьевича, выполненной на тему: XML-формат синтаксической разметки в СКАТе.

Актуальность ВКР, новизна: синтаксическая разметка является необходимой частью лингвистической аннотации корпусов текстов. Этим обуславливается актуальность работы Никиты Геннадьевича, целью которой являлось создание формата синтаксической разметки для Санкт-Петербургского корпуса агиографических текстов. Ранее, в 2012 году, уже предпринимались попытки разработать такой формат, однако этот формат оказался слишком сложным для практической реализации. Никита Геннадьевич предложил новый, упрощённый формат разметки, таким образом работа обладает новизной.

Положительные стороны ВКР: в своей работе Никите Геннадьевичу удалось разработать логичный метод представления синтаксических структур церковнославянских текстов при помощи рекомендацией консорциума TEI, при этом сохранив в рамках того же формата остальные уровни разметки, такие как морфологический и структурный. Безусловным преимуществом данной работы является то, что автор не только предложил способ представления отдельного уровня синтаксической разметки, но также создал алгоритм соотнесения уровня синтаксической разметки с уровнем структурной и морфологической разметки. Разработанная автором структура синтаксической разметки, частично основанная на синтаксическом описании, применяемом в Академической грамматике 1980 года, покрывает основные виды связей и представляет меньше сложностей при практической реализации, чем предыдущий вариант предложенной для корпуса разметки.

Практическое значение работы: разработанный автором список синтаксических отношений, а также программное обеспечение могут быть использованы в дальнейшем для работы над корпусом.

Недостатки и замечания по ВКР: к недостаткам работы относится недостаточное освещение в теоретической части работы различий между синтаксическим описанием текста в виде грамматики зависимостей и грамматики составляющих. У каждого из этих подходов есть свои преимущества и недостатки. Более подробное освещение данных аспектов позволило бы автору аргументировать его выбор в пользу грамматики зависимостей.

Хотя автор в данной работе рассматривает 10 синтаксически размеченных корпусов, в работе не упоминается такой широко используемый в настоящее время формат представления синтаксической информации как Universal Dependencies.

В качестве цели исследования заявлена синтаксическая разметка в соответствии с рекомендациями TEI, однако 9 из 10 работ, рассмотренных в теоретической части, TEI не используют, таким образом их опыт автоматически оказывается неприменимым для исследования Никиты Геннадьевича, однако автор игнорирует данный факт, не пытаясь его проанализировать. Кроме того, во многих случаях автор не объясняет в чём состоят отличия формата, используемого рассматриваемым корпусом, от рекомендаций TEI, ссылаясь на то, что читателю это должно быть очевидно.

Для создания синтаксической разметки, автор предложил использовать промежуточный формат представления в виде таблиц Excel. В рамках работы была разработана программа, которая позволяет преобразовывать разметку из этого табличного формата в формат XML, рекомендованный TEI. Однако программа не обладает возможностью пакетной работы, таким образом для преобразования каждого блока разметки требуется совершение вручную последовательности рутинных действий. Добавление пакетного режима позволило бы обрабатывать большой объём данных без участия человека.

Рекомендуемая оценка ВКР: с учётом вышеизложенного, данная работа соответствует требованиям, предъявляемым к работам данного рода и заслуживает оценки «хорошо», а автор исследования заслуживает присуждения звания магистра лингвистики.

Рецензент Алексеев Василий Анатольевич, магистр лингвистики, руководитель отдела разработок ООО «ГРАНДКАПИТАЛ».