

257/06  
03.06.2022

## ОТЗЫВ

на диссертацию Зиновьевой Анастасии Юрьевны  
«Модель многоязычного интеллектуального контент-анализа  
(на материале англо-, франко- и русскоязычных новостных сообщений  
о террористической деятельности)»,  
представленную на соискание ученой степени  
кандидата филологических наук  
по специальности 10.02.21 – Прикладная и математическая лингвистика

Диссертационное исследование Анастасии Юрьевны Зиновьевой посвящено построению модели многоязычного интеллектуального контент-анализа новостных текстов СМИ в рамках предметной области «Терроризм».

Хочется отметить высокий уровень исследовательской культуры, продемонстрированный диссертантом. В работе сочетается системный анализ исследуемого участка языковой системы, характерный для общетеоретических изысканий, и передовые методики контент-анализа.

**Актуальность** диссертационного исследования не вызывает сомнений и обусловлена направленностью на изучении важной стоящей перед лингвистикой задачи – создание методики автоматического интеллектуального анализа текстов различной тематики. Актуальности работы способствуют также выбор предметной области и стремление создать универсальную модель для работы с несколькими языками.

**Научную новизну** диссертации А.Ю. Зиновьевой мы видим в том, что в нем автор проводит анализ текстового корпуса предметной области «Терроризм» с получением релевантных для лингвистики выводов о закономерностях и ограничениях сочетаемости в соответствующем подязыке. Отдельного внимания заслуживает создание автором работы онтолексиконов для трех языков, разработка правил вывода и формирования единиц концептосферы (фреймов), организованных во взаимосвязанные иерархические структуры.

**Теоретическая ценность** исследования заключается в разработке новой универсальной модели интеллектуального контент-анализа на русском, английском и французском языках, что потребовало переосмысления ряда методологических аспектов. Полученные результаты обработки обширного текстового корпуса имеют значение и для смежных областей исследования, таких как лексикология, стилеметрия и текстометрия, лексикография.

Переходя к обзору содержания рецензируемой диссертации, отметим четкость и логичность ее композиции. Работа написана ясным языком,



положения хорошо иллюстрируются примерами и диаграммами, организованы в таблицы.

**Достоверность** выводов и положений диссертационного исследования обеспечивается корректным применением методики анализа и репрезентативным материалом, объемом более 600000 словоупотреблений, включающим три сопоставимых по объему подкорпуса текстов новостей на русском, английском и французском языках.

В первой главе А.Ю. Зиновьева ставит задачу выделить наиболее важные направления исследований по теме диссертации, а также охарактеризовать свой подход. Хорошо и полно определены основные наиболее важные понятия, использованные в работе, описаны подходы к моделированию языковых и речевых явлений, типологии контент-анализа, различных подходов к его реализации и автоматизации. Хорошее впечатление производит детальный анализ проблемы построения онтологии (с. 32 и далее) и существующих онтологий. О широкой эрудиции автора в сфере интеллектуального анализа данных и смежных областях говорит также число использованных источников, превышающее 170. При этом диссертант приводит не только обзор существующих работ, но творчески отбирает схемы и методики, оптимально отвечающие задачам исследования (с. 26, 31, 44 и далее).

Вторая и третья главы посвящены практической реализации поставленных в диссертации задач.

Следует отметить системный и взвешенный подход к организации корпусного анализа на материале новостных сообщений, принадлежащих предметной области «Терроризм». А.Ю. Зиновьева использует сочетание автоматической и ручной обработки данных (последняя неизбежна на начальном этапе работы с таким комплексным и многоаспектным объектом как речевой текст) в рамках итеративной процедуры. В результате получены цепочки лексем различной длины (*n*-граммы), что сближает работу с современными исследованиями в стилиметрии. Здесь видятся большие перспективы в плане учета распределения *n*-грамм различной длины и силы синтаксических связей для повышения точности дифференциации жанров, атрибуции, а применительно к теме рецензируемой работы – повышению точности интеллектуального анализа текста.

Совершенно оправданным является особое внимание, уделяемое глаголу (с. 65 и далее) в силу его организующей роли как структурно-семантического центра высказывания. В целом полученные данные о соотношении частотности лексем и словоупотреблений подтверждают наблюдения, сделанные в переводоведении о роли существительных как материальной основы описываемой в тексте действительности. Кроме того, данные, полученные для различных языков, ставят новые вопросы о процессе конвергенции речевых паттернов в СМИ, принадлежащих различным языковым коллективам.



Анализ репрезентации концептуальных классов в тексте выполнен корректно с учетом узкого и широкого контекста. Работа избегает достаточно распространенной ошибки, когда разные части материала анализируются на основе различных признаковых схем: здесь список концептуальных классов применяется на всем материале единообразно.

Несомненным достижением является построенная диссертантом собственная непротиворечивая онтология предметной области «Терроризм». Большое внимание уделяется логическим отношениям между элементами различного уровня, благодаря чему формируется структура упорядоченная как иерархически, так и на каждом из уровней.

Учет текстовых паттернов позволяет сформировать объективно существующие иерархии в рамках каждого из базовых концептов. Результаты представляют интерес не только для прикладной и математической лингвистики, но также для текстометрии и стилеметрии, поскольку частотность данных единиц не получила систематического рассмотрения. В возможности составления корпуса паттернов и последующего использования его для автоматической жанровой дифференциации мы усматриваем значительную **практическую** ценность работы, как и в возможности создать на основе представленной модели эффективную систему многоязычного интеллектуального контент-анализа. Более того, успешное решение поставленных задач позволило автору сделать вклад в разработку платформы концептуального аннотирования ПАНТ, включающей базу знаний и программную оболочку.

Достоинством работы является подробное описание работы программной оболочки в соответствии с алгоритмом и способы разрешения возникающих сложностей. Автор демонстрирует кропотливое и вдумчивое отношение к практической реализации сформированной модели, подробно демонстрируя ее функционирование (с.143-157).

Отметим наличие объемного раздела «Приложения» (здесь приведены текстовые репрезентации рассматриваемых концептуальных классов, иерархическая структура концептов, состав фреймов, аксиомы онтологии), который обеспечивает воспроизводимость результатов и превращает работу в ценную готовую базу для дальнейших исследований и сопоставительного анализа.

Мы можем констатировать, что все задачи, обозначенные диссертантом, успешно решены; полученные результаты полностью обоснованны; основные положения, выносимые на защиту, доказаны, а цель диссертационного исследования – достигнута.

Перспектива дальнейших исследований видится нам в расширении материала исследования путем включения других смежных предметных областей и материала на других языках, в расширении арсенала используемых методов и признаковых схемы.

Вместе с тем, хотелось бы обратить внимание на ряд дискуссионных моментов:



1. Каким образом дифференцировались близкие предметные области Терроризм и Экстремизм (см. «ультраправый экстремизм» с.187), с одной стороны, а также другие виды преступной деятельности, с другой? В частности, почему выражение «задуманный разбой» (с. 191) отнесено к репрезентантам террористической активности?
2. Зачастую организации и их деятельность признаются террористическими одними странами (и СМИ), но не признаются другими (автор сама отмечает такие случаи на с.78 и далее). О распространенности подобных случаев говорят зафиксированные в материале допускающие различную трактовку формулировки и эвфемизмы «plan to liberate the occupied territory» (с.191), «сторонник» (с.193) и др. Каким образом в этом случае осуществляется автоматический отбор материала по запросу пользователя?
3. Отмечая продуманную структуру онтологии предметной области «Терроризм», полагаем, что следовало сделать еще один шаг и экстраполировать успешно применяемую фреймовую модель на максимально высокий уровень обобщения. При этом теракт становится фреймом высшего уровня, а имеющиеся концептуальные классы – его слотами. Это позволило бы для любого интересующего текста определять его принадлежность к ядру либо периферии корпуса, а вопрос отнесения какого-либо события к акту терроризма в этом случае принял бы форму классификационной задачи с политемическими классами, для решения которой существуют различные зарекомендовавшие себя методики.
4. Работу усилило бы сопоставление данных, полученных для новостных текстов на тему «Терроризм» (частотность различных морфологических классов слов, их форм, сочетаемость и др.), с общеязыковыми данными.

В целом, однако, вышеперечисленные замечания не являются принципиальными для проведенного исследования и не сказываются на его общей положительной оценке. Они не затрагивают основных выводов диссертации, которые представляются убедительно доказанными.

Рецензируемое диссертационное исследование соответствует паспорту специальности 10.02.21 – Прикладная и математическая лингвистика. Работа актуальна по тематике, выполнена на высоком уровне, обладает научной новизной. Анализ проведен с большой тщательностью, корректно и квалифицированно. Исследование обладает как теоретической, так и практической значимостью. Автореферат и научные публикации автора отражают основное содержание диссертации. Апробация диссертации не вызывает вопросов. Список работ автора составляет 10 публикаций, 4 из них – в рецензируемых научных изданиях, включенных в Перечень ВАК



Министерства науки и высшего образования РФ, 3 – в научных изданиях, включенных в Scopus.

Все вышеизложенное позволяет заключить, что представленная к защите диссертация «Модель многоязычного интеллектуального контент-анализа (на материале англо-, франко- и русскоязычных новостных сообщений о террористической деятельности)» полностью соответствует требованиям «Положения о порядке присуждения ученых степеней» (пп. 9, 10, 11, 12, 14), утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 года №842, а ее автор, Зиновьева Анастасия Юрьевна, заслуживает присуждения ей искомой ученой степени кандидата филологических наук по специальности 10.02.21 – Прикладная и математическая лингвистика.

Доктор филологических наук, профессор,  
завкафедрой иностранных языков  
ФГБОУ ВО «Смоленский  
государственный университет»



Вадим Сергеевич Андреев

214000 Смоленск,  
ул. Пржевальского, 4  
тел.: +79156393713  
адрес электронной почты:  
[vadim.andreev@ymail.com](mailto:vadim.andreev@ymail.com)

Подпись: *Андреев В.С.*  
удостоверяю. Нач. отдела кадров СмолГУ



01 июня 2022 года