

Отзыв
официального оппонента
кандидата филологических наук Клочковой Елены Сергеевны
о диссертации Мамаева Ивана Дмитриевича «Лингвистическое
исследование скрытых сообществ в корпусе социальных медиа с
применением мультимодальных тематических моделей»,
представленной на соискание ученой степени кандидата
филологических наук по специальности 5.9.8. Теоретическая,
прикладная и сравнительно-сопоставительная лингвистика

Диссертационное исследование И.Д. Мамаева посвящено лингвистическому изучению скрытых сообществ пользователей социальных сетей методами компьютерной и корпусной лингвистики. В связи со стремительным развитием цифровой коммуникации возникла необходимость изучения и последовательного лингвистического описания речевого поведения ее участников, которое в той или иной степени, а порой и существенно, отличается от речевого поведения в реальном общении, что делает избранную тему чрезвычайно актуальной. Кроме того, актуальность работы обусловлена стремительным развитием методов обработки естественного языка, а также внедрением компьютерных и корпусных методов в анализ лингвистического материала, что ставит перед исследователями задачу совершенствования подходов и инструментария, способствующих более глубокому и достоверному описанию и моделированию языковых явлений.

Автор рецензируемой работы ставит и последовательно реализует цель, состоящую в разработке процедуры лингвистического профилирования модели скрытых сообществ, созданной методами тематического моделирования.

В качестве предмета исследования И.Д. Мамаевым избраны лингвистические параметры текстов пользователей скрытых сообществ, служащие основой создания лингвистических профилей пользователей скрытого сообщества. При этом под лингвистическим профилем понимается совокупность языковых особенностей текстов с общим тематических

компонентом, размещаемых пользователями в социальных сетях. По замыслу автора, такой подход позволяет выявить лингвистические признаки, отличающие тексты того или иного скрытого сообщества.

Таким образом, научная новизна работы заключается в разработке двух процедур компьютерного анализа корпуса текстов: процедуре выявления скрытого сообщества и построения его модели и процедуре лингвистического профилирования, позволяющей выявить релевантные лингвистические характеристики того или иного скрытого сообщества.

Теоретическая значимость исследования И.Д. Мамаева заключается в том, что она вносит вклад в дальнейшее развитие компьютерной лингвистики, методов корпусного анализа, так как содержит подробное описание и обоснование избранной процедуры анализа и алгоритма тематического моделирования, в том числе в сопоставлении с имеющимися в настоящее время альтернативами. Разработанный в работе алгоритм лингвистического профилирования может найти применение в таких областях как лингвостатистика, стилометрия, изучение жанровых особенностей текста и идиостиля автора и т.п.

Практическая ценность работы состоит в том, что разработанные в ней алгоритмы могут применяться в таких дисциплинах как социология, антропология, психология для идентификации и описания скрытых сообществ на основе создаваемых пользователями текстов.

Необходимо отметить достоверность результатов и обоснованность научных выводов, полученных И.Д. Мамаевым, что определяется следующими факторами: 1) адекватностью используемых в работе категориального аппарата и теоретической базы задачам исследования и изучаемому материалу; 2) репрезентативностью в качественном и количественном отношении того материала, который послужил основой диссертационного исследования; 3) последовательным и адекватным применением методов и инструментов обработки и моделирования лингвистических параметров.

Текст диссертации включает в себя введение, три главы, заключение и список использованной литературы, включающий 179 наименований на русском и иностранном языках.

Во введении отмечена актуальность предлагаемой работы, определены ее объект и предмет, сформулированы цели и задачи, выдвинута гипотеза, описаны методы и материал исследования. Здесь же охарактеризована актуальность и научная новизна работы, ее теоретическая и практическая значимость, представлены выносимые на защиту положения.

В главе 1 «Веб как корпус: лингвистические особенности функционирования интернет-текстов» систематизированы работы, посвященные описанию языковых и коммуникативных особенностей интернет-текстов. В главе последовательно анализируются графические, морфосинтаксические и прагмасемантические особенности интернет-текстов. В дальнейшем данный анализ служит основой для адаптации алгоритма автоматизированной предобработки текста к корпусу пользовательских постов в социальных сетях, который служит материалом исследования.

Глава 2 «Теоретические основания междисциплинарных исследований скрытых сообществ» посвящена анализу подходов к определению скрытых сообществ. На наш взгляд, особый исследовательский и практический интерес представляет обзор алгоритмов идентификации скрытых сообществ, которые включает подробный анализ трех групп алгоритмов, при этом выявляются их сильные стороны и границы применения.

В главе 3 «Разработка процедуры лингвистического профилирования скрытых сообществ» содержится описание алгоритма лингвистического профилирования, разработанного лично автором диссертации. Разработка алгоритма базируется на применении корпуса текстов пользователей социальной сети «ВКонтакте». В результате тематического моделирования и кластеризации массива текстов автор выделяет 23 скрытых сообщества и в дальнейшем выявляет значимые корреляции морфологических, синтаксических и лексических параметров в текстах каждого сообщества,

решая таким образом задачу определения характерных для каждого сообщества языковых особенностей.

В заключении работы приведены основные выводы, определяющие ее научную и методологическую ценность, а также обозначены пути дальнейшего развития исследований в данной тематической области.

Следует отметить, что научные положения и выводы, представленные в работе, подкреплены грамотным применением методов и алгоритмов компьютерной лингвистики, а также статистическими расчетами. Кроме того, наиболее важные положения иллюстрируются примерами из исследовательского корпуса, что увеличивает достоверность и убедительность рассуждений. Достоверность работы обусловлена и обширным и репрезентативным языковым материалом, составляющим порядка 10 000 постов в социальной сети «ВКонтакте», размещенных 704 пользователями.

Вопросы и замечания, возникшие в ходе чтения работы, носят дискуссионный, частный характер или формальный характер.

1. Ключевое замечание по работе состоит в следующем: несмотря на то, что работа вносит вклад в развитие компьютерной и корпусной лингвистики, а предлагаемые методики и алгоритмы могут экстраполироваться на решение других лингвистических проблем, автор недостаточно четко и явно это демонстрирует в тексте работы. Так, например, описывая теоретическую значимость работы, автор фокусируется на междисциплинарном характере работы, указывая, что разработанные в ней алгоритмы могут быть полезны в социологии, психологии и антропологии, тогда как вклад в собственно лингвистику артикулирован нечетко.
2. Еще одно замечание связано с количественными характеристиками материала. На наш взгляд, во введении следовало бы указать объем исследовательского корпуса в словоупотреблениях, что даст более внятное представление о размере корпуса.
3. На с. 47 указано, что на первом этапе создания корпуса методом сплошной выборки было сохранено более 7000 ID пользователей. Затем, после

процедуры отбора, критерием которого являлось отсутствие каких-либо дружеских связей между пользователями, в выборке осталось 704 пользователя. Как можно объяснить такую большую долю (около 90%) отсева?

4. На с. 46 (Рис. 4) в блок-схеме эксперимента указано, что этап лемматизации текстов предшествует токенизации. А на с. 48 указывается, что сначала выполнялась токенизация. Какой порядок действий следует считать верным?
5. В работе при описании лингвистических особенностей скрытых сообществ указывается на корреляцию наречий с глаголами и существительных с прилагательными и длины предложения с количеством предложных групп в этом предложении. В этой связи возникает вопрос, может ли исключаться возможность, что данные корреляции обусловлены общеязыковыми законами, а не общностью речевого поведения? Есть ли критерий, опираясь на который можно считать такую корреляцию языковой особенностью текстов конкретного скрытого сообщества, а не языковой системы как таковой?
6. Практически во всех группах отсутствует корреляция по признаку «коэффициент лексической плотности-коэффициент лексического разнообразия» (присутствует только в четырех сообществах из 23). Как можно объяснить подобное отсутствие?
7. В отдельных случаях в работе присутствуют небрежные формулировки и единичные опечатки, которые, впрочем, не мешают восприятию текста.

Высказанные вопросы и замечания носят частный характер и не снижают общего благоприятного впечатления от диссертационного исследования, направлены на его перспективу.

Работа представляет собой завершенное, оригинальное исследование, имеющее значение для изучения речевого поведения пользователей социальных сетей, а в методологическом плане для дальнейшего развития методов автоматизированного анализа языка. Работа прошла достаточную

апробацию на научных конференциях различного уровня в России и за рубежом. Автореферат диссертации и указанные в нем 7 публикаций общим объемом 4,65 п.л., в том числе 3 статьи в изданиях, рекомендованных ВАК Министерства науки и высшего образования РФ, адекватно отражают специфику исследования.

Таким образом, диссертация Ивана Дмитриевича Мамаева «Лингвистическое исследование скрытых сообществ в корпусе социальных медиа с применением мультимодальных тематических моделей» соответствует критериям пп. 9–14 Положения о присуждении ученых степеней (утв. Постановлением Правительства № 842 от 24.09.2013 г.), а ее автор заслуживает присуждения ученой степени кандидата филологических наук по специальности 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика.

Кандидат филологических наук
по специальности 10.02.04 – Германские языки
доцент кафедры иностранных языков
ФГАОУ ВО «Санкт-Петербургский
государственный электротехнический
университет «ЛЭТИ» им. В.И. Ульянова (Ленина)»

(ЕКС)

Клочкова Елена Сергеевна

5 декабря 2024 года

Контактные данные:

Федеральное государственное автономное
образовательное учреждение высшего образования
Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)

Адрес: 197022, Санкт-Петербург, ул. Профессора Попова, д. 5, лит. Ф.

Тел.: +7(812) 234-46-51

e-mail: info@etu.ru

web-сайт: <https://etu.ru>

