

## **ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 33.2.018.08,**

созданного на базе федерального государственного бюджетного образовательного учреждения высшего образования «Российский государственный педагогический университет им. А. И. Герцена» Министерства просвещения Российской Федерации **по диссертации на соискание ученой степени кандидата наук**

аттестационное дело № \_\_\_\_\_

решение диссертационного совета от 24 декабря 2024 г. № 15

о присуждении Ивану Дмитриевичу Мамаеву,  
гражданину Российской Федерации,  
ученой степени кандидата филологических наук.

Диссертация «Лингвистическое исследование скрытых сообществ в корпусе социальных медиа с применением мультимодальных тематических моделей» по специальности 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика принята к защите 8 октября 2024 года, протокол заседания № 10, диссертационным советом 33.2.018.08, созданным на базе федерального государственного бюджетного образовательного учреждения высшего образования «Российский государственный педагогический университет им. А. И. Герцена» Министерства просвещения Российской Федерации (191186, Санкт Петербург, наб. р. Мойки, 48), приказ № 406/нк от 26.04.2022 г.

Соискатель Иван Дмитриевич Мамаев, 25 октября 1996 года рождения.

В 2021 году окончил магистратуру по направлению подготовки 45.04.02 «Компьютерная и прикладная лингвистика» в федеральном государственном бюджетном образовательном учреждении высшего образования «Санкт-Петербургский государственный университет», присвоена квалификация «Магистр».

В период с 2021 по 2024 гг. обучался в аспирантуре федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет», освоил программу подготовки научно-педагогических кадров в аспирантуре по направлению подготовки 45.06.01 Языкознание и литературоведение, присвоена квалификация «Исследователь. Преподаватель-исследователь».

В период подготовки диссертации и по настоящее время соискатель работает старшим преподавателем кафедры Р7 «Теоретической и прикладной лингвистики» федерального государственного бюджетного образовательного учреждения высшего образования «Балтийский государственный технический университет "ВОЕНМЕХ" им. Д. Ф. Устинова».

С 2022 года работает ассистентом на кафедре математической лингвистики федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» (по совместительству).

Работа выполнена на кафедре математической лингвистики федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет» Правительства Российской Федерации.

**НАУЧНЫЙ РУКОВОДИТЕЛЬ** – Митрофанова Ольга Александровна, кандидат филологических наук, доцент, доцент кафедры математической лингвистики федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет».

**ОФИЦИАЛЬНЫЕ ОППОНЕНТЫ:**

Колмогорова Анастасия Владимировна, доктор филологических наук, профессор, профессор Департамента филологии Санкт-Петербургской Школы гуманитарных наук и искусств, заведующий лабораторией языковой конвергенции федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский

университет «Высшая школа экономики» в Санкт-Петербурге;

Клочкова Елена Сергеевна, кандидат филологических наук, доцент кафедры иностранных языков федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)» – дали положительные отзывы на диссертацию.

**ВЕДУЩАЯ ОРГАНИЗАЦИЯ:** федеральное государственное автономное образовательное учреждение высшего образования «Южно-Уральский государственный университет (национальный исследовательский университет)», г. Челябинск в своем положительном отзыве, подписанном кандидатом филологических наук, доцентом, заведующим кафедрой лингвистики и перевода Института лингвистики и международных коммуникаций Бабиной Ольгой Ивановной и кандидатом педагогических наук, доцентом, директором Института лингвистики и международных коммуникаций Ярославовой Еленой Николаевной, указала, что актуальность исследования И. Д. Мамаева определяется необходимостью изучения скрытых сообществ и глубокого понимания динамики взаимодействия в них, поскольку анализ таких сообществ может помочь в выявлении суицидальных наклонностей и криминальных и экстремистских ячеек. Научная новизна диссертации состоит в том, что впервые разработана процедура выявления скрытых сообществ путем применения методов машинного обучения для тематического моделирования сообщений пользователей социальной сети. Разработка критериев сбора материала и методов построения корпуса текстов для моделирования скрытых сообществ, а также предложенная процедура лингвистического профилирования на основе корреляционного анализа обладают теоретической значимостью и вносят вклад в развитие компьютерной и математической лингвистики. Проведенное исследование имеет междисциплинарный потенциал, поскольку внедрение

лингвистического анализа в существующие методики выделения групп пользователей вносит вклад в исследования скрытых сообществ в социологии, психологии, антропологии, криминологии и других областях. Практическая значимость исследования заключается в разработке методик, которые могут стать основой для проведения исследований и практической деятельности в сфере социальных наук. Отмечается убедительность и достоверность полученных результатов.

Соискатель имеет 7 опубликованных работ по теме диссертации, общим объемом 4,65 п.л., авторский вклад – 3,68 п.л., из них в рецензируемых научных изданиях опубликовано 3 работы (2,07 п.л./1.78 п.л.), статей в других изданиях – 4 (2,58 п.л./1.9 п.л.). В диссертации отсутствуют недостоверные сведения об опубликованных соискателем ученой степени работах.

**Наиболее значимые работы по теме диссертации:**

- 1. Мамаев И. Д. Лингвистические параметры для идентификации скрытых сетевых сообществ / И. Д. Мамаев, О. А. Митрофанова // Terra Linguistica. – 2024. — Т. 15. — №. 1. — С. 102-115. (0,96/0,67 п.л.)**
- 2. Мамаев И. Д. Лингвистические профили скрытых сообществ: морфосинтаксический аспект / И. Д. Мамаев // Филологические науки. Вопросы теории и практики. — 2024. — Т. 17. — Вып. 4. — С. 1155-1162. (0,57 п.л.)**
- 3. Мамаев И. Д. Кластерный анализ лингвистических профилей скрытых сообществ / И. Д. Мамаев // Филологические науки. Вопросы теории и практики. — 2024. – Т. 17. — Вып. 5. — С. 1739-1747. (0,54 п.л.)**

На диссертацию и автореферат поступило 4 положительных отзыва.

Отзывы поступили:

1. От доктора технических наук, профессора, профессора кафедры прикладной информатики и информационной безопасности Высшей школы кибертехнологий, математики и статистики федерального государственного

бюджетного образовательного учреждения высшего образования «Российский экономический университет им. Г. В. Плеханова» Чеповского Андрея Михайловича.

Содержит замечания:

В работе активно используется термин «корреляция», но нет однозначного определения что именно автор понимает под данной характеристикой, которая может иметь разные смыслы и способы вычисления.

Нет четкого разделения понятий сообщества в социальном графе, как графе взаимодействующих объектов, к которому применяются методы анализа топологии графа, и визуализации облака слов, или облака текстов.

2. От кандидата технических наук, доцента, доцента высшей школы лингвистики и педагогики федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого» Коган Марины Самуиловны.

Содержит вопросы:

Делалась ли попытка оценить количество пользователей, «публикующих посты на большое количество общих тем» (с.13)? Как решался вопрос с отнесением их к определенной тематической группе? Были ли случаи с отнесением таких пользователей в несколько групп?

Какова максимальная и средняя численность выявленных сообществ? Для определения лингвистического профиля важен размер постов, написанных пользователями, принадлежащих конкретному скрытому сообществу. Определялся ли средний объем таких подкорпусов или границы его изменения?

3. От кандидата филологических наук, доцента, доцента кафедры русского и иностранных языков федерального казенного образовательного учреждения высшего образования «Вологодский институт права и экономики Федеральной службы исполнения наказаний» Воробьевой Елены Николаевны.

Вопросов и замечаний нет.

4. От доктора филологических наук, профессора кафедры русского языка, современной русской и зарубежной литературы федерального

государственного бюджетного образовательного учреждения высшего образования «Воронежский государственный педагогический университет» Литвиновой Татьяны Александровны.

Содержит вопросы:

Чем обусловлено использование сервиса Profiling-UD при наличии многочисленных библиотек для автоматического анализа текстов? Проверяли ли Вы точность полученных на его основе результатов?

Насколько уместно говорить о лексической гомогенности текстов соцсетей, коль скоро была использована всего одна метрика для исследования данного уровня (с. 15) - лексическая корреляция «коэффициент лексической плотности - коэффициент лексического разнообразия»? Отмечу также, что в данном случае неясным является вкладываемый автором работы смысл термина «корреляция». Необходимо пояснить утверждение о том, что (с. 20) «лексические параметры практически не представлены в лингвистических профилях».

Поступившие отзывы аффилированы в соответствии с требованиями пункта 28 Положения о порядке присуждения ученых степеней.

**Выбор официальных оппонентов и ведущей организации обосновывается их компетентностью и достижениями в соответствующей отрасли науки, наличием публикаций в соответствующей сфере исследования, а также решением диссертационного совета 33.2.018.08 от 8 октября 2024 года, протокол № 10 в соответствии с пунктами 22, 24 «Положения о присуждении ученых степеней», утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842 (в действующей редакции).**

**Диссертационный совет отмечает, что на основании выполненных соискателем исследований:**

**предложены процедуры выявления скрытых сообществ, построения их модели и лингвистического профилирования;**

**проведена адаптация алгоритма автоматической предобработки текста к**

корпусу пользовательских постов социальной сети «ВКонтакте»;

**исследованы** лингвостатистические характеристики текстов скрытых сообществ, позволяющие описать лингвистический профиль каждого из выявленных сообществ;

**доказана** перспективность тематического моделирования и кластеризации корпуса для выделения скрытых сообществ;

**выявлена** языковая специфика текстов, представляющих скрытые сообщества;

**разработаны** критерии для построения корпуса текстов для решения задачи выявления скрытых сообществ.

**Теоретическая значимость исследования обоснована тем, что:**

**разработана** новая научная идея о том, что алгоритмы тематического моделирования позволяют выявлять сообщества людей, имеющих схожие интересы и публикующих посты в социальных сетях;

**доказана** обоснованность разработанных процедур анализа и алгоритма тематического моделирования;

**предложены** алгоритм лингвистического профилирования скрытых сообществ и система лингвостатистических метрик для их сравнения;

**проведен** системный сопоставительный анализ трех групп алгоритмов лингвистической идентификации скрытых сообществ с установлением их сильных сторон и границ применения;

**доказана** эффективность процедуры выявления скрытых сообществ посредством применения методов машинного обучения и математической статистики.

**Значение полученных соискателем результатов исследования для практики подтверждается тем, что:**

**доказано**, что предложенные методики являются основой для проведения исследований и практической деятельности в сфере прикладной лингвистики, социальной информатики, социологии и других смежных областях гуманитарного знания;

**разработанные** алгоритмы и процедуры рекомендуются к использованию в качестве методологической основы для исследования других корпусов текстов;

**подтверждена** перспективность использования разработанного перечня критериев отбора текстов социальных сетей для последующего выявления так называемых скрытых сообществ;

**представлены** результаты, допускающие внедрение разработанного алгоритма лингвистического профилирования в сервисы социальных сетей для модерации сообществ, мониторинга общественного мнения и других задач, требующих извлечения и анализа контента, используемого пользователями сетей.

**Оценка достоверности результатов исследования выявила:**

**что теория построена** на основе многоаспектного подхода к анализу речевых продуктов, входящих в исследовательский корпус, включающего современные положения теории языка, корпусной лингвистики и квантитативной лингвистики;

**идея базируется на** применении корпусного подхода, ориентированного на практическое изучение языка в реальных ситуациях и материалах, а также инструментария лингвостатистики;

**использованы** методологический инструментарий корпусной лингвистики; современные методики сбора и обработки исходной информации, позволившие включить в исследовательский корпус не менее 10 000 постов в социальной сети «ВКонтакте», размещенных 704 пользователями;

**применен** широкий набор программно-инструментальных средств моделирования.

**Личный вклад соискателя состоит в:**

- научной лингвистической интерпретации процедур тематического моделирования и лингвистического профилирования для решения задачи выделения скрытых сообществ;

- разработке и апробации алгоритма лингвистического профилирования, позволившего выявить 23 скрытых сообщества;
- определении графических, морфосинтаксических и прагмасемантических особенностей интернет-текстов;
- выявлении значимых корреляций морфологических, синтаксических и лексических параметров в текстах каждого сообщества;
- разработке собственной дискретной автор-тематической модели, позволяющей на основе тематических микромоделей каждого автора с приписанными вручную метками построить тематическую модель корпуса.

В ходе защиты диссертации были высказаны следующие замечания: в работе недостаточно артикулирован вклад проведенного исследования в собственно лингвистику; не указан объем исследовательского корпуса в словоупотреблениях; необходимо введение дефиниции термина мультимодальность; неясно, какой граф – ориентированный или неориентированный – представляет собой адекватную модель семантической сети для выявления скрытых сообществ; недостаточно обоснована связь между оцениваемой лексической корреляцией и выводом о гомогенности текстов постов; не определено, насколько данные корреляции обусловлены общеязыковыми законами, а не общностью речевого поведения; осталось неясным «отсеивание» более 90% из 7000 пользователей сети и трактовка уровня гранулярности захвата пользователей тематическими моделями.

Соискатель И. Д. Мамаев согласился с замечаниями, ответил на заданные ему в ходе заседания вопросы и привел собственную аргументацию: в работе применены методы корреляционного анализа, выявлены особенности речевого поведения русскоязычных пользователей социальных сетей, предложенный подход может быть применен для изучения социолингвистических различий между группами говорящих; объем корпуса составляет более 1.5 млн словоупотреблений; в тематическом моделировании *мультимодальность* определяется как наличие нескольких видов корпусных метаданных, учитываемых при построении модели; в корпусе постов

отсутствуют данные об имплицитном влиянии или взаимном воздействии пользователей друг на друга, поэтому адекватной моделью семантической сети для выявления скрытых сообществ является неориентированный граф; отсутствие лексической корреляции во многих лингвистических профилях связано с тем, что в текстовых выборках небольшого объема статистические показатели не отражают устойчивых языковых закономерностей, в этой связи делается вывод о гомогенности исследуемых текстов; исследуемые корреляции могут встречаться в моделях речевого поведения, но значимость корреляций в скрытых сообществах подтверждает языковую специфику выделенных групп, проводится сравнение с другими постами того же объема и жанра, в которых такие корреляции не имеют значимости; отбор пользователей проводился по критерию отсутствия дружеских связей, что уменьшило влияние социальных факторов на модель скрытых сообществ, гранулярность данных зависела от выбора постов объемом более 200 словоупотреблений и использования стоп-словаря, что важно для точности тематического моделирования.

**На заседании 24 декабря 2024 г. диссертационный совет пришел к выводу о том, что диссертация Ивана Дмитриевича Мамаева представляет собой научно-квалификационную работу, которая соответствует критериям, предъявляемым к диссертациям на соискание ученой степени кандидата наук (п.9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства РФ от 24.09.2013 № 842, в действующей редакции), и принял решение: за решение научной задачи разработки методологии выявления скрытых сообществ с помощью применения к их текстовой продукции алгоритмов тематического моделирования и лингвистического профилирования присудить Ивану Дмитриевичу Мамаеву **ученую степень кандидата филологических наук.****

При проведении тайного голосования диссертационный совет в количестве 14 человек, из них 7 докторов наук по научной специальности

рассматриваемой диссертации, участвовавших в заседании, из 15 человек, входящих в состав совета, проголосовали: за – 13, против – 1, недействительных бюллетеней нет.

ПРЕДСЕДАТЕЛЬ



Беляева Лариса Николаевна  
доктор филологических наук,  
профессор

УЧЕНЫЙ СЕКРЕТАРЬ

Кампилова Ольга Николаевна  
кандидат филологических  
наук, доцент

Дата оформления заключения «24» декабря 2024 года.