МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Лемаев Владислав Игоревич

Кросс-языковое распознавание эмоций в тексте и речи: лингвистические аспекты

Специальность 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата филологических наук

Москва – 2026

Диссертация подготовлена на кафедре теоретической и прикладной лингвистики филологического факультета Московского государственного университета имени М.В. Ломоносова

Научный руководитель: **Лукашевич Наталья Валентиновна**
доктор технических наук, профессор

Официальные оппоненты: **Колмогорова Анастасия Владимировна**
доктор филологических наук, профессор, Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербургская школа гуманитарных наук и искусств, департамент филологии, профессор

Литвинова Татьяна Александровна
доктор филологических наук, Воронежский государственный педагогический университет, гуманитарный факультет, кафедра русского языка, современной русской и зарубежной литературы, профессор

Бабина Ольга Ивановна
кандидат филологических наук, доцент, Южно-Уральский государственный университет, институт лингвистики и международных коммуникаций, кафедра «Лингвистика и перевод», заведующий кафедрой

Защита диссертации состоится «17» сентября 2026 года в 16:00 часов на заседании диссертационного совета МГУ.058.1 Московского государственного университета имени М.В. Ломоносова по адресу: 119991, Москва, ГСП-1, Ленинские горы, д. 1, стр. 51, 1-й учебный корпус, филологический факультет, ауд. 1060.

E-mail: sovet@philol.msu.ru

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский проспект д. 27) и на сайте ДС МГУ: https://dissovet.msu.ru/dissertation/3945

Автореферат разослан «___» _____ 2026 г.

Учёный секретарь
диссертационного совета
кандидат филологических наук И.В. Одинцова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Настоящая работа посвящена изучению влияния лингвистических признаков, таких как базовый порядок слов, пересечение лексики и используемые обучающие данные, на качество межъязыкового переноса обученных языковых моделей в задаче автоматического распознавания эмоций.

Автоматическое распознавание эмоций как в тексте, так и в речи является важным направлением в области обработки естественного языка (англ. *Natural Language Processing*), позволяя повысить качество взаимодействия человека с компьютером. В то время как для английского языка имеется большое количество размеченных наборов данных, или датасетов, для анализа эмоций, доступные данные для большинства языков, включая русский, часто имеют ограниченный объём и недостаточно высокое качество разметки. В случае недостатка обучающих данных часто используется межъязыковой перенос языковых моделей (англ. *cross-lingual transfer*), при котором языковые модели изначально обучаются на данных высокоресурсных языков, а затем уже применяются к данным малоресурсных языков. Одной из слабоизученных сторон данного подхода по-прежнему остаётся влияние свойств самих языков, использованных в процессе переноса. В частности, данная проблема ранее совсем не рассматривалась в рамках задачи автоматического распознавания эмоций.

Таким образом, **актуальность** данной работы обусловлена несколькими факторами. Во-первых, разработка в области распознавания эмоций и развитие диалоговых систем для взаимодействия человека и компьютера крайне востребованы и требуют дальнейших исследований. Во-вторых, для многих языков наблюдается недостаток качественно размеченных эмоциональных данных, в связи с чем необходимо более эффективное обучение языковых моделей на уже имеющихся данных, прежде всего на данных других языков. В-третьих, влияние сходства языков, использованных в процессе обучения и

межъязыкового переноса языковых моделей, в задаче автоматического распознавания эмоций по-прежнему остаётся слабоизученным.

Степень научной разработанности проблемы. Задача автоматического распознавания эмоций рассматривается в работах П. Экмана, Р.Р. Пикар, К.Р. Шерера, Ф. Делларта, Р. Коуи, И. Вагнера, В. Бобичев, М. Мициоса, Р. Михалчи, Д. Багата, Д.Е. Кахиани, Х.М. Файка, Г. Онвуджекве, Д. Ли, Х. Ху, У. Дейв, П. Джафарзаде, Т. Нве, М. Шами, К. Гобла, Р. Банзе и других, а также в рамках серии соревнований SemEval. Межъязыковой перенос языковых моделей и мультиязычные модели рассматривались в работах М. Пикулиака, Ф. Филиппи, А. Конно, Г. Лампла, А. Сринивасан, Дж. Пфайффера, Х. Хуанга, З. Хана и других. Влияние сходства используемых языков на результаты межъязыкового переноса языковых моделей в различных задачах обработки естественного языка рассмотрены в работах А. Дешпанде, З. Ву, А. Лаушер, С.Г. Упадхай, Ю.С. Лин, С.М. Ферару, Ч.Л. Лю, Ж. Ву, К. Ахуджи, В. Патил, Т. Пиреса и других.

Объект исследования – межъязыковой перенос обученных языковых моделей для решения задачи автоматического распознавания эмоций.

Предмет исследования – влияние лингвистических признаков на качество межъязыкового переноса обученных языковых моделей в задаче автоматического распознавания эмоций в письменных текстах и устной речи.

Цель исследования – установить влияние лингвистических признаков, таких как базовый порядок слов, пересечение лексики и используемые обучающие данные, на качество межъязыкового переноса обученных языковых моделей в задаче автоматического распознавания эмоций на основе современных подходов и доступных обучающих данных. Первоочередное внимание при этом обращается на данные русского языка, который выступает целевым языком при межъязыковом переносе моделей, предварительно обученных на других языках.

Для достижения данной цели были поставлены следующие **задачи**:

1) определить современное состояние области автоматического распознавания эмоций, текущие достижения в ней, а также применяемые подходы для решения этой задачи;
2) определить современное состояние области автоматического распознавания эмоций для русского языка, имеющиеся для него обучающие данные и доступные модели, установить их качество путём экспериментального исследования;
3) определить наиболее значимые лингвистические признаки, влияющие на качество межъязыкового переноса языковых моделей в задачах обработки естественного языка, и установить степень их влияния на межъязыковой перенос языковых моделей, обученных решению задачи автоматического распознавания эмоций. В качестве экспериментального материала использовать текстовые и речевые данные русского языка, выступающего в качестве целевого языка для переноса.

В настоящем исследовании выдвигается следующая **гипотеза**: лингвистические признаки языка-источника и целевого языка: генетическая и фонологическая близость, сходство синтаксиса, пересечение лексики и используемые для обучения и тестирования модели языковые данные существенно влияют на качество межъязыкового переноса моделей, обученных решению задачи автоматического распознавания эмоций; при этом чем сходство языков выше, тем выше должно быть и качество переноса.

В работе использовались следующие **методы исследования**:

1) общенаучные методы: описание, анализ, классификация, индукция и дедукция, эксперимент, сравнение;
2) лингвистические методы: описательный, сопоставительный, количественный, лексико-семантический, корпусный;
3) лингвокультурологический метод (анализ культурных особенностей, свойственных отдельным языкам, и как они выражаются в лексике).

В качестве **теоретической базы** послужили работы в следующих областях:

1) типология языков и сопоставительное языкознание – работы М.С. Драйера, М. Хаспельмата, Дж. Литтелла, Д.Р. Мортенсена, С. Вихманна, Э.В. Холмана, Г. Беллы, К. Батсурэна, Ф. Джункильи;
2) способы выражения эмоций в разных языках – работы Ф. Филиппи, Дж. Джексона, К.А. Линдквист, Дж. Уоттса, Т.Р. Генри, П. Экмана, В.В. Фризена, С. Анколи, Л. Кайзера, Р. Банзе, С.Г. Упадхай, С.М. Ферару;
3) создание эмоциональных датасетов и корпусов – работы Дж. А. Миллера, Э. Камбрии, Н.В. Лукашевич, Ж. Коссаифи, И.С. Энгберг, К. Буссо, М. Незами, С.Р. Ливингстона, Д. Демски, Ш. Х. Мухаммеда, Р. Лотиана, В. Кондратенко, К. Чжоу;
4) алгоритмы и подходы для решения задачи распознавания эмоций – работы Т. Миколова, А. Васвани, Н. Шазира, Н. Пармар, Э.Л. Мааса, М. Петерса, Х. Ху, Р. Коуи, Ф. Алотаиби, Р. Михалчи, Д. Багата, Х.М. Файка, Д. Ли, У. Дейв, М.К.Х. Ли, М.З. Бойто, Ю. Лю, М. Отта.

Научная новизна настоящего диссертационного исследования заключается в следующем:

1) выявлено влияние лингвистических признаков на качество межъязыкового переноса языковых моделей в задаче автоматического распознавания эмоций в устной речи;
2) выявлено влияние лингвистических признаков на качество межъязыкового переноса языковых моделей в задаче автоматического распознавания эмоций в письменных текстах;
3) установлено, что лингвистические признаки, влияющие на качество межъязыкового переноса моделей распознавания эмоций, для текста и речи различаются.

Теоретическая значимость исследования состоит в описании возможностей современных языковых моделей при решении задачи автоматического распознавания эмоций на данных русского языка и переноса

на русский язык языковых моделей, обученных на других языках. Были описаны лингвистические признаки, оказывающие влияние на качество межъязыкового переноса моделей в задаче распознавания эмоций, такие как обучающие данные, лексика и синтаксис, а также выделены наиболее значимые из них. Сформулированы проблемы, из-за которых снижается качество межъязыкового переноса моделей в задаче автоматического распознавания эмоций, включая описание особенностей имеющихся в открытом доступе обучающих данных для русского и других языков.

Практическая значимость диссертационной работы выражается в описании влияния лингвистических признаков и методов сбора обучающих данных на качество межъязыкового переноса моделей в задаче автоматического распознавания эмоций, а также выявлении проблем, наблюдающихся в доступных на данный момент русских датасетах. Полученные в результате экспериментов данные можно использовать для улучшения уже имеющихся русских корпусов эмоциональных высказываний, а также при создании новых корпусов. Описанное влияние лингвистических признаков может быть также использовано при создании и разметки новых корпусов и при решении реальных задач межъязыкового переноса моделей для распознавания эмоций, прежде всего на данных русского языка, но потенциально применимо и к данным других языков.

Экспериментальным материалом диссертационного исследования послужили русский датасет эмоциональной речи Dusha, русский датасет эмоциональных текстов CEDR, английский датасет эмоциональной речи IEMOCAP, польский датасет эмоциональной речи nEMO, китайский датасет эмоциональной речи ESD, японский датасет эмоциональной речи JVNV, мультиязычный датасет эмоциональных текстов BRIGHTER, в частности данные русского, немецкого, испанского, китайского, хинди и татарского языков.

На защиту выносятся следующие **положения**:

1. Качество переноса речевых моделей для автоматического распознавания эмоций повышается при обучении моделей на языке, близком к целевому. Качество переноса повышается ещё сильнее при обучении моделей сразу на нескольких близких языках.
2. Высокое качество переноса речевых моделей для автоматического распознавания эмоций достигается, в частности, за счёт сходства интонации высказываний и, как следствие, сходства актёрских навыков говорящих. Сходство в лексике, семантике и синтаксисе языков значимого влияния на качество переноса речевых моделей не оказывает.
3. Повышение качества переноса текстовых моделей для автоматического распознавания эмоций при их обучении на языке, близком к целевому, достигается за счёт сходства способов выражения эмоций в языке-источнике и целевом языке. Качество переноса моделей повышается, когда в данных обоих языков для выражения одних и тех же эмоций используются схожие маркеры: лексика, эмодзи и знаки препинания.
4. Лексика является главным фактором, на основе которого языковые модели распознают эмоции в письменных текстах, при этом на качество межъязыкового переноса больше всего влияет сходство в семантике. Чем меньше контекстов имеют схожие по семантике эмоциональные слова, чем более они однозначны, тем выше качество переноса языковых моделей. В то же время схожесть написания слов и разные типы письменности в используемых языках не оказывают значительного влияния на качество переноса.
5. Невербальные элементы, такие как знаки препинания и эмодзи, по сравнению с лексикой имеют второстепенное значение и меньший вес при классификации эмоций в письменных текстах. Влияние невербальных элементов больше всего проявляется в случае, когда

большое их количество встречается в обучающих данных только с одной эмоцией.

6. Как и в случае речевых моделей, сходство синтаксиса и, прежде всего, сходство базового порядка слов не оказывают значимого влияния на качество переноса текстовых моделей для автоматического распознавания эмоций.

Достоверность результатов настоящей диссертационной работы обеспечена использованием современных и общепринятых подходов обработки языка, подробным описанием использованных методов и открытом доступе к полученным результатам для возможности их сравнения и воспроизведения, а также публикацией основных результатов диссертационной работы в рецензируемых научных журналах, что подтверждает их апробацию в научном сообществе.

Апробация работы и публикации. Результаты исследования опубликованы в 5 статьях в изданиях, рекомендованных для защиты в диссертационном совете МГУ им. М.В. Ломоносова по специальности 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика (филологические науки). Основные положения докладывались на семинаре «Машинное обучение в автоматической обработке текстов» НИВЦ МГУ имени М.В. Ломоносова.

Личный вклад соискателя заключается в проведении аналитического обзора современных методов автоматического распознавания эмоций на основе моделей-трансформеров, проведением основного объёма практических исследований и описанием их результатов под руководством научного руководителя, где вклад соискателя был решающим.

Структура. Работа состоит из введения, четырёх глав, заключения, списка литературы и двух приложений. Общий объём работы составляет 152 страницы. Список литературы содержит 128 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** обоснованы цель и актуальность исследования, его задачи, теоретическая и практическая значимость, использованный экспериментальный материал, а также приводятся основные положения, выносимые на защиту.

В **первой главе** диссертации «Теоретические аспекты автоматической классификации эмоций» приводятся история области автоматического распознавания эмоций, подробный обзор стандартного инвентаря эмоций, используемого в задаче их распознавания, описание типов датасетов для обучения языковых моделей и методов их составления как для речевых, так и для текстовых данных, а также приводятся описание используемых подходов для решения задачи автоматического распознавания эмоций и примеры их реального применения. В **параграфе 1.1** описаны теоретические и практические положения, используемые для выделения набора эмоций. В качестве основных подходов применяют классификацию данных либо на основе определённого небольшого набора эмоций (около 6 штук), либо на основе двухмерного пространства «интенсивность/валентность». В **параграфе 1.2** приведена краткая история области автоматического распознавания эмоций, указаны ключевые научные работы, способствовавшие её развитию. В **параграфе 1.3** описаны основные типы датасетов и методы их создания в случае речевых и в случае текстовых данных, приведены их примеры. Показаны дополнительные факторы, влияющие на качество эмоциональных датасетов. В **параграфе 1.4** рассмотрены методы предобработки данных, применяемые перед подачей данных на вход алгоритму распознавания эмоций. Описан процесс предобработки текстов (токенизация, лемматизация, векторизация), а также описаны способы выделения признаков в речевых данных. В **параграфе 1.5** приведён разбор алгоритмов классификации, используемых для решения задачи автоматического распознавания эмоций, а также приведены примеры используемых на их основе подходов.

Во **второй главе** диссертации «Межъязыковой перенос моделей» приведено описание подходов к межъязыковому переносу обученных языковых моделей. Описаны цели, для достижения которых применяется межъязыковой перенос, главной из которых является решение проблемы недостатка обучающих данных для малоресурсных языков. Описаны проблемы, возникающие при обучении мультиязычных моделей, главная из которых – снижение качества работы модели при её обучении на слишком большом количестве языков. В **параграфе 2.1** описаны основные лингвистические признаки языков, которые, как было показано в различных научных работах, наибольшим образом влияют на качество межъязыкового переноса моделей. Среди них выделяют структурное сходство языков (синтаксис, порядок слов, близость языков), пересечение лексики и качество используемых обучающих данных. В **параграфе 2.2** кратко описано влияние параметров самой языковой модели на качество переноса, в частности влияние её архитектуры и параметров обучения.

В **третьей главе** диссертации «Распознавание эмоций для русского языка» рассмотрено текущее положение дел в области автоматического распознавания эмоций на русском языке. Описаны последние исследования на эту тему, применяющиеся подходы и имеющиеся в открытом доступе данные. На данное время в русском языке имеется определённое количество размеченных данных, однако по-прежнему наблюдаются ограниченность охватываемых ими предметных областей и сравнительно небольшое их количество, особенно текстовых данных.

В **параграфе 3.1** для оценки качества современных методов автоматического распознавания эмоций на данных русской разговорной речи проведено экспериментальное исследование с использованием датасета Dusha, являющегося в настоящее время самым большим датасетом русской эмоциональной речи. В качестве моделей для тестирования выбраны две актуальные архитектуры для обработки речи на основе трансформеров: HuBERT и WavLM, которые показали высокую эффективность в смежных

задачах обработки речи. Для них использовались две конфигурации base и large, содержащие около 95 миллионов и 315 миллионов параметров соответственно.

Были использованы общедоступные предобученные модели, дополнительно обученные на соответствующем датасете. Обучение и тестирование производились на всём объёме обучающих данных. В качестве метрики оценки качества их работы использовалась WA (Weighted Accuracy), которая учитывает соотношение количества примеров для каждой эмоции при расчёте оценки. В табл. 1 представлены полученные результаты:

Таблица 1 – Результаты обученных на русском языке речевых моделей для распознавания эмоций

Модель	Гнев	Позитив	Грусть	Нейтральное состояние	Другое	WA
HuBERT_base	0.83	0.78	0.79	0.94	0.71	0.866
HuBERT_large	0.87	0.81	0.80	0.93	0.7	0.874
WavLM_base	0.87	0.77	0.83	0.94	0.82	0.872
WavLM_large	0.86	0.81	0.84	0.93	0.75	**0.878**

Наилучшие результаты показала модель WavLM_large (0.878); с небольшим отставанием второе место заняла HuBERT-large (0.874). Тем самым архитектура WavLM демонстрирует лучшее качество по сравнению с архитектурой HuBERT как при сравнении конфигураций base, так и при сравнении конфигураций large. При этом модель HuBERT с конфигурацией large всё ещё показывает результаты лучше, чем модель WavLM с конфигурацией base, что говорит о том, что размерность и количество параметров модели имеют значительное влияние на качество работы даже при учёте различий в архитектуре.

Отдельные эмоции распознаются хорошо, лучше всего классифицируются гнев и нейтральное состояние. Наибольшие трудности возникли при распознавании класса «другое». Данная группа слишком обширна, поэтому необходима более детальная категоризация эмоций внутри неё.

Качество самих данных также может вызывать трудности при классификации: датасет содержит множество примеров, которые озвучены относительно нейтрально и вызывают проблемы с классификацией даже у носителей языка. Кроме того, в данных часто присутствует посторонний шум.

В **параграфе 3.2** представлено экспериментальное тестирование современных языковых моделей на текстовых эмоциональных данных. Для этого использованы модели RuBERT и rubert-tiny2 – две русские предобученные модели на основе архитектуры BERT, отличающиеся преимущественно только количеством параметров (rubert-tiny2 является уменьшенной версией RuBERT). Также была протестирована мультиязычная модель XLM-RoBERTa, созданная на основе архитектуры RoBERTa. Обучение и тестирование проводились на русском датасете эмоциональных текстов CEDR и русских эмоциональных данных из датасета BRIGHTER. В качестве метрики использовалась макро-усреднённая F-мера. Полученные результаты представлены в табл. 2:

Таблица 2 – Результаты обученных на русском языке текстовых моделей для распознавания эмоций

Датасет	Модель	Гнев	Страх	Радость	Грусть	Удивление	Отвращение	Общее качество
CEDR	RuBERT	0.79	0.82	0.84	0.79	0.9	-	0.83
	rubert-tiny2	0.78	0.83	0.85	0.83	0.87	-	0.83
	XLM-RoBERTa	0.85	0.83	0.87	0.83	0.78	-	0.83
BRIGHTER	RuBERT	0.82	0.74	0.95	0.8	0.79	0.77	0.81
	rubert-tiny2	0.84	0.76	0.93	0.88	0.84	0.81	0.84
	XLM-RoBERTa	0.86	0.82	0.96	0.88	0.84	0.8	**0.86**

Полученные результаты демонстрируют сопоставимо высокое качество работы всех моделей на обоих датасетах, наилучшее качество показывает

мультиязычная модель. Качество классификации отдельных эмоций у всех моделей также схоже.

В **четвёртой главе** диссертации «Эксперименты по межъязыковому переносу моделей для задачи классификации эмоций» представлены основные экспериментальные исследования данной работы по межъязыковому переносу языковых моделей, обученных для решения задачи автоматического распознавания эмоций на речевых и текстовых данных.

В **параграфе 4.1** представлено исследование межъязыкового переноса речевых моделей на русский язык при их изначальном обучении на иностранных языках-источниках. Для исследования качества межъязыкового переноса речевых моделей использована модель HuBERT-base, имеющая около 95 млн параметров. Использовалась именно конфигурация base, а не large, для более точного сравнения с готовой мультиязычной моделью с тем же количеством параметров. Обучение проводилось отдельно на 4 языках-источниках (английском, польском, китайском и японском), после чего обученные модели были протестированы на данных русского языка. Русские эмоциональные данные во время обучения не использовались.

Обучающие данные взяты из 4 иностранных датасетов эмоциональной речи: английского IEMOCAP, польского nEMO, китайского ESD и японского JVNV. Тестирование проводилось на русском датасете Dusha. Для классификации использовались 3 эмоции: гнев, радость и грусть, которые были выбраны на основе их встречаемости и достаточного количества примеров во всех упомянутых датасетах. Для каждого иностранного датасета составлена обучающая выборка из 600 аудиозаписей по 200 на каждую эмоцию. Тестовая выборка с таким же объёмом составлена и для русского датасета Dusha. Всего для каждого языка-источника было обучено 5 моделей с разными параметрами обучения:

1) base – модель обучалась с нуля только на выборке из 600 примеров из соответствующего датасета;

2) full – модель обучалась на всех имеющихся данных и классах эмоций соответствующего датасета, после чего дополнительно обучалась на соответствующей выборке из 600 примеров (для её адаптации под распознавание только трёх эмоций);
3) pretrain – схожа с конфигурацией full, но обучение происходило не с нуля, а использовалась модель, уже предобученная на английском речевом корпусе LibriSpeech объёмом около 960 часов. Во время обучения все слои модели были размморожены;
4) pretrain_freeze – повторяет конфигурацию pretrain, за исключением того, что во время обучения были заморожены все слои, кроме последних двух (целью было выяснить, как близость языков влияет на процесс обучения модели);
5) multi – повторяет конфигурацию pretrain, но в качестве предобученной модели использовалась мультиязычная модель mHuBERT-147, предобученная на речевых данных объёмом более 90 тысяч часов из 147 языков (включая английский, польский, китайский, японский и русский).

Полученные результаты классификации для всех моделей приведены в табл. 3. В качестве метрики качества классификации использовалась UA (Unweighted Accuracy). Для подсчёта качества классификации каждой отдельной эмоции использовалась F-мера. К названию каждой модели в начало добавлен префикс, обозначающий язык обучения.

Таблица 3 – Результаты переноса речевых моделей для распознавания эмоций на русский язык

Модель	Качество классификации			
	Гнев	Радость	Грусть	Общее качество
eng-base	0.46	0.28	0.51	0.42
pol-base	0.48	0.33	0.54	**0.46**
chi-base	0.46	0.16	0.60	0.46
jap-base	0.47	0.34	0.47	0.43
eng-full	0.52	0.03	0.44	0.41
pol-full	0.38	0.48	0.53	**0.46**
chi-full	0.12	0.35	0.59	0.43
jap-full	0.44	0.18	0.56	0.43
eng-pretrain_freeze	0.58	0.35	0.37	0.45
pol-pretrain_freeze	0.54	0.41	0.63	**0.56**
chi-pretrain_freeze	0.16	0.41	0.57	0.45
jap-pretrain_freeze	0.48	0.08	0.26	0.35
eng-pretrain	0.65	0.25	0.57	0.52
pol-pretrain	0.60	0.40	0.65	**0.58**
chi-pretrain	0.26	0.17	0.55	0.42
jap-pretrain	0.49	0.11	0.26	0.45
eng-multi	0.63	0.24	0.73	0.59
pol-multi	0.58	0.68	0.73	**0.68**
chi-multi	0.24	0.33	0.58	0.46
jap-multi	0.54	0.09	0.67	0.52

Основные выводы на основе полученных результатов:

1) Близость языка-источника и целевого языка положительно влияет на качество переноса моделей. Наилучшее качество переноса для всех конфигураций модели наблюдается при обучении на польском языке, который ближе всех использованных языков к русскому.
2) Мультиязычная модель показывает наилучший результат, при этом качество её работы также выше, если её дообучение решению задачи распознавания эмоций происходит на языке, наиболее близком к русскому (см. более высокое качество у моделей eng-multi и pol-multi, по сравнению с аналогичными моделями на китайском и японском языках).
3) При использовании обучающих данных нескольких языков их близость к языку-источнику также положительно влияет на качество переноса. У конфигураций pretrain и pretrain_freeze, уже предварительно обученных

на большом объёме английской речи, дополнительное обучение на английских и польских эмоциональных данных показывает повышение качества классификации эмоций по сравнению с конфигурациями base и full, не имеющих такого же предварительного обучения. При этом качество повышается ещё сильнее при разморозке и обучению всех параметров модели, как в конфигурации pretrain. У аналогичных моделей, дообученных на китайском и японском языках, более дальних русскому, качество классификации либо снижается, либо почти не меняется.

4) Важным фактором распознавания эмоций в речи является интонация, а лексика и семантика почти не оказывают на это влияния. Самый яркий этому показатель – противопоставление радости/гнева и грусти на основе интенсивности интонации, где грусть выражается с низкой интенсивностью и классифицируется стабильно хорошо, а у моделей возникают проблемы с различением между собой высокоинтенсивных гнева и радости.
5) Интонация также влияет на качество переноса языковых моделей для распознавания эмоций в речи – сильное влияние оказывает сходство самих данных разных языков и методов их сбора. Например, японский датасет JVNV был записан с привлечением профессиональных актёров, и большинство примеров содержат экспрессивную речь, а в записи датасета Dusha участвовали случайные люди, что выражается в частой монотонной речи в примерах, – это ведёт к тому, что модель jap-full классифицирует около 60% всех тестовых русских примеров как грустные из-за их низкой интенсивности.
6) Значимого влияния синтаксического сходства языков, в частности базового порядка слов, на качество переноса речевых моделей для распознавания эмоций не наблюдается. Все использованные языки имеют базовый порядок SVO, в то время как японский имеет базовый порядок SOV. При этом сбалансированные по количеству обучающих данных модели pol-base и jap-base показывают сравнительно схожие результаты.

Отдельные примеры тоже классифицируются сравнительно одинаково, например высказывание “*а мне нравится ход ваших мыслей*” (порядок SVO), произнесённое с грустной интонацией, и польской, и японской моделью неверно классифицируется с эмоцией гнева.

7) Использование большого объёма несбалансированных данных и дополнительных эмоций при обучении, как в конфигурации full, может значительно повлиять на качество классификации отдельных эмоций в речи. Например, у моделей pol-full и chi-full по сравнению с моделями pol-base и chi-base наблюдается улучшение качества классификации эмоции радости при падении качества классификации эмоции гнева, что может быть объяснено наличием в соответствующих полных датасетах (nEMO и ESD) эмоции удивления, которая в речи часто выражается сходно с эмоцией радости. У модели eng-full, наоборот, радость почти не распознаётся, так как в полном датасете IEMOCAP имеется дополнительная негативная эмоция раздражения, на которую приходится около трети всех обучающих примеров датасета.

В **параграфе 4.2** представлено исследование межъязыкового переноса текстовых моделей на русский язык при их изначальном обучении на иностранных языках-источниках. Для этого на русский язык были перенесены модели, обученные на эмоциональных текстовых данных немецкого, испанского, китайского, хинди и татарского языков. Для классификации использовалась модель XLM-RoBERTa. Все данные взяты из мультиязычного эмоционального датасета BRIGHTER, который использовался в рамках прошлогоднего соревнования SemEval-2025 Task 11, основной задачей которого был межъязыковой перенос текстовых моделей для распознавания эмоций. Датасет BRIGHTER содержит тексты, извлечённые из постов в социальных сетях и классифицированные на 6 эмоций: гнев, отвращение, страх, радость, грусть, удивление. Русские эмоциональные данные во время обучения не использовались. Для всех использованных в исследовании языков

представлено подробное описание имеющихся для них данных, а также способы выражения отдельных эмоций, их сходства и различия.

Из-за относительно небольшого количества обучающих для использованных языков, особенно для отдельных эмоций (например, всего 56 примеров с эмоцией отвращения в китайском), было трудно собрать сбалансированные выборки для всех 6 классов эмоций, поэтому подход к эксперименту был иным по сравнению с речевыми данными. Для каждого языка модель обучалась только в одной конфигурации на всём имеющемся объёме обучающих данных (такой подход показал наилучшие результаты во время тестирования). В качестве меры оценки работы моделей использовалась взвешенная F-мера, которая учитывает дисбаланс количества примеров в разных классах. Полученные результаты межъязыкового переноса моделей представлены в табл. 4:

Таблица 4 – Результаты переноса текстовых моделей для распознавания эмоций на русский язык

Эмоция	Язык обучения				
	Немецкий	Испанский	Китайский	Хинди	Татарский
Гнев	0.57	0.56	0.57	0.57	0.55
Отвращение	0.26	0.2	0	0.08	0.11
Страх	0.43	0.81	0	0.71	0.27
Радость	0.68	0.77	0.69	0.75	0.68
Грусть	0.09	0.69	0.39	0.39	0.68
Удивление	0	0.65	0	0.56	0.47
Общее качество	0.43	**0.63**	0.39	0.55	0.51

Для подтверждения того, что на полученные результаты не влияет несбалансированность использованных обучающих выборок, в работе представлены результаты работы моделей при их обучении на очень небольших, но сбалансированных выборках. Распределение качества классификации отдельных эмоций у моделей в таком случае остаётся схожим, что подтверждает достоверность полученных в табл. 4 результатов.

Кроме того, были также проведены дополнительные эксперименты для исследования влияния на качество переноса моделей схожести синтаксиса

между языком-источником и целевым языком переноса, а также значимости лексики и невербальных элементов (например, знаков препинания) для распознавания эмоций.

Для исследования влияния синтаксического сходства языков на качество переноса моделей проведено сравнение качества классификации эмоций в зависимости от базового порядка слов (взаимное расположение субъекта, объекта и глагола) в исходном языке обучения и порядка слов в русских тестовых примерах. Для всех использованных в работе языков приведено описание их базового порядка слов. После этого приведено тестирование качества классификации моделей в зависимости от порядка слов в русских тестовых примерах, а также подсчитана статистическая значимость полученных результатов на основе пермутационного статистического теста.

Для дополнительного исследования значимости лексики и невербальных элементов было проведено сравнение качества классификации русских тестовых примеров, во-первых, при наличии и отсутствии в них отрицания, а во-вторых, при наличии и отсутствии в них определённых знаков препинания (восклицательный и вопросительный знаки, многоточие). Для всех полученных результатов подсчитана статистическая значимость на основе пермутационного статистического теста.

Основные выводы о межъязыковом переносе текстовых моделей для распознавания эмоций (в примерах, выделенных курсивом, сохранены орфография и пунктуация источника):

1) Наилучшее качество классификации наблюдается при обучении модели на данных испанского языка. Для остальных языков качество классификации тоже неплохое, независимо от письменности (на качество переноса не влияют иероглифическое письмо в китайском и деванагари в хинди).
2) В случае малого количества обучающих примеров для отдельных эмоций они не распознаются вовсе (например, китайская модель не распознаёт эмоции отвращения, страха и удивления).

3) На качество классификации эмоций очень сильно влияет схожесть самих данных в языке-источнике и целевом языке. Так, гнев и радость стабильно хорошо распознаются для всех языков, потому что во всех них они выражаются схоже. Но это не так, например, для грусти: в русских данных она выражается прежде всего за счёт эмодзи («*эх,ну так нельзя(((*») и эксплицитной лексики, – этот способ также используется в испанских и татарских данных, для которых грусть распознаётся хорошо. У других языков грусть выражается имплицитно через контекст, без частого использования эмодзи: например, предложение «*никаких постоянных увлечений нет,совсем.*» правильно распознаётся как грустное моделью, обученной на данных хинди, но не распознаётся испанской и татарской моделями.
4) Синтаксические различия в языках, в частности базовый порядок слов, не оказывают значимого влияния на качество переноса текстовых моделей, причём это касается как влияния на общее качество классификации, так и на качество классификации отдельных эмоций. Однако в данном случае исследование проведено на относительно небольшом объёме данных, – вполне возможно, что при достаточно большом их количестве результат может оказаться другим. При этом в предсказаниях моделей всё же наблюдается придание большего веса отдельным позициям в тексте, в частности финальной позиции (например, «*ублюдки бесят меня на протяжении всего дня!!!!*» правильно классифицируется как гневное только 2 моделями из 5, тогда как «*всё, этот говнюк меня раззадорил. меня ещё никто так не бесил.*» правильно классифицируется 4 моделями, имея тот же глагол «бесить» в конце текста).
5) Сходство лексики – главный фактор при классификации, при этом важна не схожесть написания, а семантика и встречаемость в

однотипных контекстах. Так, пример «*они у меня милашки*» все модели распознают верно независимо от используемой в языке-источнике письменности. В то же время эмоция отвращения распознаётся плохо для всех языков из-за того, что она выражается исключительно через специфическую лексику, которая во отдельных случаях может также использоваться для выражения эмоции гнева, – именно гнев в большинстве случаев модели и предсказывают для русских примеров (например, для «*просыпаться в полдвенадцатого - отвратительно*» все модели предсказывают эмоцию гнева).

6) Элементы отрицательной полярности (слова «не», «нет», «никто» и т.д.) для всех языков показывают значимую корреляцию с эмоцией гнева, что подтверждается статистическим тестом. При этом наблюдаются повышение качества классификации эмоции гнева и снижение качества классификации радости при наличии в тексте отрицания: например, «*как хорошо не ходить в шк*» содержит положительно окрашенное слово «хорошо» и неправильно классифицируется всеми моделями либо с эмоцией гнева, либо с эмоцией грусти, – если убрать частицу «не», то все модели правильно предсказывают радость, что указывает на потенциально больший вес у отрицания при предсказании, чем у остальной лексики.
7) Невербальные маркеры, такие как эмодзи и знаки пунктуации, являются второстепенными по сравнению с лексикой и имеют меньший вес при предсказании эмоций. Кроме того, эмодзи одинаково хорошо распознаются как при обучении на испанском языке, в котором эмодзи выражаются графически (например, «:)»), так и при обучении на татарском, в котором эмодзи, как и в русском, выражаются через последовательные круглые скобки (например, «(((»). Однако в отдельных случаях невербальные маркеры могут значительно влиять на качество классификации, например при

наличии их большого количества только у одной эмоции в обучающих данных, что нужно учитывать при их сборе (например, в обучающих данных на татарском языке вопросительный знак преимущественно встречается только в текстах с эмоцией удивления, из-за чего имеет очень сильную корреляцию с этой эмоцией и большой вес при предсказании).

8) В языковых моделях наблюдается наличие внутреннего пространства для классов эмоций, другими словами, в процессе обучения модели также усваивают отношения отдельных эмоций между собой, а не рассматривают их как независимые категории. Свидетельством этому служат примеры с многоточием, которое у большинства моделей коррелирует с эмоцией грусти: в примере «*...очень страшно...*» эмоция страха выражается эксплицитно, но многоточие имеет больший вес и 4 из 5 моделей приписывают эмоцию грусти, а радостный пример «*сидела на работе...а тут раз звонок, встреча, цветы...*» и гневный пример «*эти наушники неудобные... дерьмо*» правильно классифицируются 4 из 5 моделей, – в них лексика имеет больший вес, чем многоточие. Эмоция страха в использованных обучающих данных для всех языков выражается с низкой интенсивностью, а эмоции радости и гнева, напротив, выражаются с высокой интенсивностью, что указывает на противопоставление этих эмоций внутри модели на основе категории интенсивности.

ЗАКЛЮЧЕНИЕ

Полученные в ходе настоящего исследования результаты подтверждают влияние близости языков на качество межъязыкового переноса моделей распознавания эмоций. Было установлено, что на качество межъязыкового переноса речевых и текстовых моделей влияют разные лингвистические признаки.

Анализ работы речевых моделей показал, что важным признаком для распознавания эмоций в речи служит интонация, тогда как значимого влияния лексического и синтаксического сходства используемых языков не наблюдается. Наилучшие результаты показали модели, обученные на языках, близких к целевому. Использование дополнительных обучающих данных из нескольких близких языков также повышает качество переноса.

Исследование текстовых моделей, в свою очередь, показало, что главным фактором, влияющим на качество переноса, является схожесть выражения эмоций в обучающих и целевых данных, а качество классификации выше всего тогда, когда в них совпадают способы выражения эмоций. Дополнительные исследования продемонстрировали, что наиболее значимый фактор при классификации эмоций – это лексика, а невербальные маркеры второстепенны, но всё ещё сильно влияют на предсказания моделей в зависимости от контекста и их встречаемости в обучающих данных. Текстовые модели хорошо распознают лексику независимо от различий в письменности, при этом наиболее важно совпадение семантического значения слов, а не их написание. Значимого влияния базового порядка слов на качество переноса моделей не наблюдается, тем не менее результаты показывают, что модели придают больше веса отдельным позициям в тексте, в частности финальной позиции.

Для повышения устойчивости результатов переноса и речевых, и текстовых моделей распознавания эмоций необходимы систематически собранные, сбалансированные и нормализованные датасеты, в идеале специально созданные именно для этой задачи, так как на качество переноса

влияют и качество собранных данных, и методы их сбора, и набор эмоций, используемый в датасете.

Направлением дальнейшего исследования представляется сбор большего объёма размеченных данных для анализа возможностей переноса языковых моделей распознавания эмоций, а также использование данных большего числа языков. Другим потенциальным направлением является применение более комплексных подходов и моделей, включая мультиязычный подход с одновременным использованием и речевых, и текстовых данных.

Список работ, опубликованных по теме диссертации

Научные статьи, опубликованные в изданиях, рекомендованных для защиты в диссертационном совете Московского государственного университета имени М.В. Ломоносова по специальности 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика (филологические науки)

1. *Лемаев В.И., Лукашевич Н.В.* Автоматическая классификация эмоций в речи: методы и данные // Litera. 2024. – № 4. – С. 159–173. EDN: WOBSMN. Импакт-фактор журнала 0,333 (РИНЦ). 0,93 п.л. Объём авторского вклада: 9/10.
2. *Лемаев В.И., Лукашевич Н.В.* Межъязыковой перенос моделей автоматического распознавания эмоций в речи на русский язык: данные и тенденции // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2025. – № 5. – С. 28–38. EDN: ATUXZG. Импакт-фактор журнала 0,833 (РИНЦ). 0,68 п.л. Объём авторского вклада: 9/10.
3. *Лемаев В.И., Лукашевич Н.В.* Выражение эмоций в разных языках: возможности межъязыкового переноса моделей распознавания // Вестник Московского университета. Серия 9: Филология. 2025. – № 4. – С. 9–20. EDN: MOXNLD. Импакт-фактор журнала 0,288 (РИНЦ). 0,75 п.л. Объём авторского вклада: 9/10.
4. *Лемаев В.И., Лукашевич Н.В.* Влияние порядка слов на качество межъязыкового переноса моделей для распознавания эмоций на русском языке // Rhema. Рема. 2025. – № 4. – С. 53–77. EDN: SNYKGQ. Импакт-фактор журнала 0,636 (РИНЦ). 1,5 п.л. Объём авторского вклада: 9/10.
5. *Лемаев В.И.* Влияние лексики и пунктуации на качество межъязыкового переноса текстовых языковых моделей распознавания эмоций // Litera. 2026. – № 3. – С. 20–44. EDN: HSXJFK. Импакт-фактор журнала 0,333 (РИНЦ). 1,5 п.л.