

В.И. Лемаев, Н.В. Лукашевич

Московский государственный университет
имени М.В. Ломоносова,
119991 г. Москва, Российская Федерация

Влияние порядка слов на качество межъязыкового переноса моделей для распознавания эмоций на русском языке

В данной работе рассмотрен перенос обученных языковых моделей с одного языка на другой для задачи автоматической классификации эмоций в тексте и кратко описаны основные современные подходы для решения этой задачи. Основной целью было исследование влияния порядка слов на качество классификации перенесенных моделей на основе данных соревнования SemEval-2025 Task 11. Для исследования была использована модель XLM-RoBERTa, отдельно обученная на немецком, испанском, китайском, хинди и татарском языках и затем протестированная на данных русского языка. Для полученных результатов были проведены анализ и подсчет их статистической значимости, на основе которых были описаны основные факторы, влияющие на качество переноса языковых моделей.

Ключевые слова: обработка естественного языка, межъязыковой перенос, распознавание эмоций, трансформеры, XLM-RoBERTa

Для ЦИТИРОВАНИЯ: Лемаев В.И., Лукашевич Н.В. Влияние порядка слов на качество межъязыкового переноса моделей для распознавания эмоций на русском языке // Рема. Rhema. 2025. № 4. С. 53–77. DOI: 10.31862/2500-2953-2025-4-53-77



DOI: 10.31862/2500-2953-2025-4-53-77

V. Lemaev, N. LoukashevitchLomonosov Moscow State University,
Moscow, 119991, Russian Federation

The influence of word order on the quality of cross-linguistic transfer of emotion recognition models in Russian

This study explores the transfer of pretrained language models from one language to another for the task of automatic emotion classification in text and provides a brief overview of current state-of-the-art approaches to this problem. The primary objective was to investigate the impact of word order on the classification performance of transferred models using the dataset from the SemEval-2025 Task 11 competition. For the experiments, the XLM-RoBERTa model was employed, separately fine-tuned on German, Spanish, Chinese, Hindi, and Tatar, and subsequently evaluated on Russian-language data. The obtained results were analyzed for statistical significance, allowing the identification of key factors influencing the effectiveness of cross-lingual transfer in language models.

Key words: natural language processing, cross-lingual transfer, emotion recognition, transformers, XLM-RoBERTa

FOR CITATION: Lemaev V., Loukashevitch N. The influence of word order on the quality of cross-linguistic transfer of emotion recognition models in Russian. *Rhema*. 2025. No. 4. Pp. 53–77. DOI: 10.31862/2500-2953-2025-4-53-77

1. Введение

Автоматическое распознавание эмоций является одной из популярных задач в сфере обработки естественного языка и состоит в приписывании всему тексту или его отрывку определенной эмоции. На практике алгоритмы распознавания эмоций могут применяться как для улучшения популярных в настоящее время виртуальных собеседников (чат-ботов), позволяя им выстраивать более естественный и понятный

диалог с человеком, так и для более целенаправленных задач, например, для постановки диагноза в области здравоохранения [Hansen et al., 2022] или для обеспечения общественной безопасности [Lefter, Jonker, 2017].

Современные языковые модели, такие как BERT и GPT, предобучались на большом объеме обучающих данных, в которых превалирует английский язык. При этом многие языки мира имеют ограниченный объем данных, доступный для обучения языковых моделей. В таких условиях альтернативой часто выступает перенос на целевой язык моделей, заранее обученных на другом языке, но данный подход пока не достиг тех же результатов, на которые способен человек [Pikuliak, 2021]. Применимо к задаче автоматического распознавания эмоций межъязыковой перенос также осложняет разница в подходах исследователей к составлению датасетов, в том числе разница в используемых наборах эмоций.

Целью данной работы является исследование влияния порядка слов в предложении (в первую очередь влияние базового порядка слов, т.е. взаимного расположения субъекта, объекта и глагола) на качество классификации эмоций в тексте при переносе текстовых языковых моделей с одного языка на другой на основе реальных данных, используемых на практике.

В исследовании использовалась мультязычная модель XLM-RoBERTa [Conneau et al., 2020], дообученная на эмоциональных текстах немецкого, испанского, китайского, хинди и татарского языков и затем протестированная на русском языке. При обучении и тестировании были использованы данные из датасета BRIGHTER [Muhammad et al., 2025a], собранные в рамках соревнования SemEval-2025 Task 11 [Muhammad et al., 2025b]. В соответствии с правилами подзадания Track C, которое в данном соревновании состояло в межъязыковом переносе моделей, обученных для решения задачи распознавания эмоций, в текущем исследовании при обучении модели также не были использованы данные целевого (русского) языка (подход zero-shot).

Структура данной работы оформлена следующим образом: в разделе 2 представлен разбор предыдущих работ и описание лингвистических факторов и их влияния на качество классификации при переносе обученной языковой модели с одного языка на другой. В разделах 3 и 4 содержатся краткое описание применяемых в исследованиях наборов эмоций и современных подходов к классификации эмоций соответственно. В разделе 5 содержится описание использованных в данной работе обучающих и тестовых данных, а также языковой модели. Наконец, в разделе 6 представлены полученные в рамках данного исследования результаты и их описание.

2. Влияние лингвистических факторов на качество переноса обученных языковых моделей

На данный момент существует достаточное количество исследований о влиянии лингвистических признаков на качество результатов работы обученных языковых моделей при их межъязыковом переносе. Однако большинство таких исследований рассматривает только наиболее общие задачи обработки языка, такие как распознавание именованных сущностей, частеречная разметка, ответы на вопросы и т.д., почти не затрагивая задачу распознавания эмоций. Так, в работе [Philippy et al., 2023] собрано подробное описание факторов, влияющих на качество переноса языковых моделей. Всего авторы выделяют пять таких факторов: структурное сходство языков, пересечение лексики между ними, данные для обучения, архитектура языковой модели и гиперпараметры ее обучения. В настоящей работе в первую очередь рассматривается влияние лингвистических факторов, поэтому мы не будем подробно останавливаться на архитектуре моделей и их гиперпараметрах, а сосредоточимся только на первых трех факторах.

К *структурному сходству* языков в первую очередь относятся *синтаксис* и *порядок слов*. Хотя они традиционно считаются одними из самых важных признаков, исследования показывают, что при переносе языковых моделей они не всегда оказывают значимое влияние. Например, в работах [Deshpande et al., 2022; Wu et al., 2023] авторы проводят сравнение между случайными перестановками слов и целенаправленной адаптацией порядка слов под другой язык. Как было обнаружено, если в предложениях первого языка порядок слов адаптирован под порядок слов второго языка, то качество классификации языковой модели при ее переносе снижается меньше, чем в случае случайного порядка слов в первом языке.

Среди основных лингвистических признаков также выделяют *близость языков*, в частности, генетическую, географическую и фонологическую. Согласно исследованиям [Lin et al., 2019; Lauscher et al., 2020], все типы близости значительно влияют на качество классификации при межъязыковом переносе языковых моделей, при этом фонологическая близость является более значимым критерием для задач на уровне обработки токенов (например, для распознавания именованных сущностей), чем для задач на уровне обработки всего предложения. Определение структурного сходства языков обычно происходит на основе типологических баз данных, например, WALIS [Dryer, Haspelmath, 2013], и векторов лингвистических признаков языков – чаще всего для этого применяется библиотека lang2vec [Littell et al., 2017].

Под *пересечением лексики* языков подразумевается доля слов, присутствующих в обоих исследуемых языках, которая рассчитывается как процент общих для двух языков слов от суммы всех уникальных слов, встречающихся в этих языках. Для подсчета пересекающейся лексики также часто используется расстояние Левенштейна [Wichmann et al., 2010] и метрика ezGlott [Kovacevic et al., 2022]. Как показывают исследования, хотя пересечение лексики и оказывает влияние на качество межъязыкового переноса языковых моделей, это влияние почти никогда не является изолированным. Так, в работе [Patil et al., 2022] на материале близких языков было установлено, что большее количество пересекающейся лексики значительно сильнее улучшает качество переноса в тех случаях, когда количество обучающих данных исходного языка мало. Работа [Deshpande et al., 2022] же демонстрирует, что пересечение лексики значительнее всего влияет на качество работы перенесенной модели тогда, когда порядки слов у языков различаются.

Влияние *обучающих данных* на качество межъязыкового переноса моделей в первую очередь зависит от *размера обучающего корпуса* исходного языка. Так, увеличение количества обучающих данных на этапе предобучения значительно улучшает качество переноса современных языковых моделей на основе трансформеров, таких как mBERT [Devlin et al. 2019], особенно в сравнении с более старыми нейросетевыми подходами, у которых рост качества при увеличении объема обучающих данных в определенный момент замедлялся вплоть до остановки [Liu et al., 2020].

Помимо этого, для мультиязычных моделей на этапе предобучения также важно *соответствие между корпусами* всех используемых языков, причем в первую очередь более важна принадлежность корпусов к одной и той же области, чем их параллельность [Deshpande et al., 2022].

3. Составление инвентаря эмоций

Одной из частых проблем, возникающих при переносе обученных языковых моделей, является различие в подходах к классификации данных. Для области автоматического распознавания эмоций эта проблема стоит особенно остро, т.к. эмоции имеют высокую вариативность и значимые различия в их выражении в зависимости от языка и культурного бэкграунда говорящего [Jackson et al., 2019]. Типичные наборы эмоций, используемые в различных областях психологии, могут насчитывать около 300 разнообразных эмоций и их вариаций. Однако для задач автоматического распознавания исследователи обычно выделяют 6 эмоций, которые являются универсальными для большинства языков и называются

«архетипическими»: это гнев, отвращение, страх, радость, грусть и удивление [Cowie et al., 2001]. Эти 6 эмоций и образуют базовый набор, используемый в большинстве исследований.

Другой подход, который часто применяется наравне с небольшим набором эмоций, – это классификация данных на основе двумерного пространства. В качестве первого измерения используется понятие «возбуждение», которое описывает степень интенсивности выражаемой эмоции. Возбуждение часто коррелирует с физиологическими процессами, происходящими в теле человека при выражении определенных эмоций. Например, повышенный сердечный ритм и быстрая речь характерны для эмоций гнева, страха и радости, в то время как, наоборот, пониженный сердечный ритм и медленная речь характерны для эмоции грусти. В текстах степень возбуждения можно часто оценить по наречиям, например, *очень* и *слегка*.

В качестве второго измерения выступает понятие «валентность», отражающее степень положительности или, наоборот, отрицательности выражаемой эмоции. Валентность помогает дифференцировать эмоции гнева и радости, для которых часто характерна одинаково высокая степень интенсивности, и в текстах валентность часто соотносится с положительно или отрицательно окрашенными словами. Преимуществом подхода на основе возбуждения/валентности является то, что обученная модель получает более четкое представление о соотношении эмоций между собой и даже в случае ошибки может с большей вероятностью предсказать эмоцию, близкую к истинной [Mitsios et al., 2024].

4. Подходы к классификации

Наиболее популярными в настоящее время являются подходы на основе самообучения, позволяющие алгоритмам самостоятельно находить закономерности в исходных данных и тем самым обрабатывать большие объемы данных с минимальной предварительной экспертной разметкой. Это стало возможно в первую очередь благодаря появлению архитектуры трансформера [Vaswani et al., 2017], в основе которой лежит модифицированная идея механизма внимания, состоящая в том, что при решении конкретной задачи языковая модель может обращать больше внимания на определенные части данных в зависимости от контекста и требований самой задачи. Данная модель смогла не только улучшить предыдущие результаты в различных задачах, но и также стала способна одновременно обрабатывать большее количество данных, тем самым ускорив этап обучения. На архитектуре трансформера базируются как большие языковые модели (англ. LLM), например, GPT [Radford et al.,

2018], так и более классические BERT [Devlin et al., 2019] и RoBERTa [Liu et al., 2019]. В ходе соревнования SemEval-2025 Task 11, на котором были поставлены задачи как автоматического распознавания эмоций предложений в рамках одного языка, так и переноса обученных моделей на другие языки, наилучшие результаты показали подходы на основе ансамблей LLM-моделей и написанных для них промт-запросов.

В рамках данной статьи для задачи переноса используется мультиязычная модель XLM-RoBERTa-base¹, обученная на текстовых данных 100 языков общим объемом 2,5 терабайт, представленных в открытой библиотеке CommonCrawl² (большинство текстов написаны на английском языке). Модель XLM-RoBERTa основана на модели RoBERTa, представляющей собой оптимизированную архитектуру BERT, которая улучшила качество ее работы за счет обучения на большем объеме данных с увеличенным количеством шагов самого обучения, а также других небольших изменений. Предобучение модели XLM-RoBERTa производилось на задаче MLM (англ. Masked Language Modeling), состоящей в предсказании предварительно замаскированных в предложении слов за счет других слов, стоящих по соседству.

5. Данные и модель

В качестве данных был взят мультиязычный датасет BRIGHTER [Muhammad et al., 2025a], содержащий письменные эмоциональные тексты для 28 языков. В качестве источников для данного датасета выступили посты в социальных сетях, личные блоги, новостные источники и вручную или автоматически сгенерированные предложения. В датасете представлены 6 эмоций для каждого языка: гнев/anger, отвращение/disgust, страх/fear, радость/joy, грусть/sadness, удивление/surprise. Для каждого языка тексты внутри датасета разбиты на обучающую, валидационную и тестовую выборки. Из-за небольшого объема имеющихся данных, а также для воспроизводимости результатов исследования, в данной работе использовались исходные выборки без удаления или добавления дополнительных примеров за исключением следующих случаев:

- 1) в данной работе использовались только те примеры, которые выражают как минимум одну эмоцию;
- 2) в случае выражения нескольких эмоций одновременно истинной была принята первая по списку.

¹ URL: <https://huggingface.co/FacebookAI/xlm-roberta-base>

² URL: <https://commoncrawl.org/>

В качестве алгоритма классификации была использована модель XLM-RoBERTa-base. В рамках текущей работы данная модель была отдельно дообучена на данных каждого исходного языка, представленного в датасете BRIGHTER (использовались данные обучающей и валидационной выборок), и затем перенесена и протестирована на тестовой выборке русского языка без дополнительного обучения. В качестве основной метрики качества классификации во всех примерах использовалась взвешенная F-мера. На основе результатов переноса модели было отобрано 5 языков с наилучшим общим качеством классификации: немецкий, испанский, китайский, хинди и татарский.

В качестве зависимой переменной для исследования выступил базовый порядок слов, в частности, порядки SVO, SOV и VSO. Рассмотрим базовый порядок слов в данных языках более подробно, приведенные сведения опираются на работу [Dryer, Haspelmath, 2013].

Немецкий: имеет комплексную систему порядка слов, которая выражается в использовании порядка V2 (verb second) в главной клаузе – при этом в таких случаях преобладает порядок SVO – и порядка SOV в зависимых клаузах. В то же время, несмотря на предпочтительное расположение S перед O в предложении, каких-то реальных ограничений на это нет, и их порядок является относительно свободным.

Испанский: базовым порядком считается порядок SVO. В то же время испанский имеет достаточно богатую морфологию, что на практике выражается в относительно свободном порядке слов.

Китайский: базовым порядком является порядок SVO.

Хинди: базовым порядком официально считается порядок SOV, при этом порядок может меняться с целью изменения прагматического смысла высказывания.

Татарский: базовым порядком слов является порядок SOV.

Русский: базовым порядком слов традиционно считается порядок SVO, но при этом, благодаря богатой морфологии, на практике порядок слов является в значительной степени свободным.

В таблице 1 кратко приведены данные о базовом порядке слов для каждого из упомянутых языков.

В таблице 2 для каждого языка представлено распределение обучающих данных по эмоциям, в таблице 3 – общее количество примеров в русской тестовой выборке. Объем валидационных выборок для каждого языка составил около 15% от объема соответствующей обучающей выборки (около 10% для хинди и около 20% для татарского языка): использовались валидационные выборки, представленные в датасете.

Таблица 1

**Базовый порядок слов в рассматриваемых
в данной статье языках**
[Basic word order in the languages used in this paper]

Язык [Language]	Порядок слов [Word order]
Немецкий [German]	V2 в главной клаузе, SOV в зависимой клаузе [V2 in the main clause, SOV in the subordinate clause]
Испанский [Spanish]	SVO
Китайский [Chinese]	SVO
Хинди [Hindi]	SOV
Татарский [Tatar]	SOV
Русский [Russian]	SVO

Таблица 2

**Количество обучающих примеров для каждой эмоции
в разных языках**
**[Number of training examples for each emotion
in each language used for training]**

Эмоция [Emotion]	Язык [Language]				
	Немецкий [German]	Испанский [Spanish]	Китайский [Chinese]	Хинди [Hindi]	Татарский [Tatar]
Гнев [Anger]	768	492	1178	422	137
Отвращение [Disgust]	294	257	56	181	108
Страх [Fear]	159	295	46	369	48
Радость [Joy]	481	579	523	442	182
Грусть [Sadness]	201	192	189	346	214
Удивление [Surprise]	55	181	47	191	126
Всего [Total]	1958	1996	2039	2556	815

Таблица 3

**Количество примеров в русской тестовой выборке
для каждой эмоции**
[Number of examples in the Russian test dataset for each emotion]

Эмоция [Emotion]	Количество [Number of examples]
Гнев [Anger]	452
Отвращение [Disgust]	172
Страх [Fear]	212
Радость [Joy]	386
Грусть [Sadness]	206
Удивление [Surprise]	144
Всего [Total]	1572

В таблицах 4–8 для каждого языка, данные которого использовались для обучения модели, указано количество обучающих примеров для каждого порядка слов в зависимости от эмоции.

Таблица 4

**Количество примеров для каждой эмоции
в немецкой обучающей выборке в зависимости
от порядка слов в предложении**
[Number of examples for each emotion in the German training dataset
by the word order in the sentence]

Эмоция [Emotion]	Порядок слов [Word order]		
	SVO	SOV	VSO
Гнев [Anger]	221	147	50
Отвращение [Disgust]	79	51	22
Страх [Fear]	29	39	11
Радость [Joy]	107	85	33
Грусть [Sadness]	44	36	12
Удивление [Surprise]	19	9	3
Всего [Total]	499	367	131

Таблица 5

**Количество примеров для каждой эмоции
в испанской обучающей выборке в зависимости
от порядка слов в предложении**
[Number of examples for each emotion in the Spanish training dataset
by the word order in the sentence]

Эмоция [Emotion]	Порядок слов [Word order]		
	SVO	SOV	VSO
Гнев [Anger]	86	52	22
Отвращение [Disgust]	34	26	6
Страх [Fear]	48	43	13
Радость [Joy]	69	63	20
Грусть [Sadness]	29	24	11
Удивление [Surprise]	16	26	7
Всего [Total]	282	234	79

Таблица 6

**Количество примеров для каждой эмоции
в китайской обучающей выборке в зависимости
от порядка слов в предложении**
[Number of examples for each emotion in the Chinese training dataset
by the word order in the sentence]

Эмоция [Emotion]	Порядок слов [Word order]		
	SVO	SOV	VSO
Гнев [Anger]	534	11	193
Отвращение [Disgust]	30	0	3
Страх [Fear]	22	0	15
Радость [Joy]	225	1	85
Грусть [Sadness]	79	3	29
Удивление [Surprise]	17	0	12
Всего [Total]	907	15	337

Таблица 7

**Количество примеров для каждой эмоции
в обучающей выборке на языке хинди
в зависимости от порядка слов в предложении**
[Number of examples for each emotion in the Hindi training dataset
by the word order in the sentence]

Эмоция [Emotion]	Порядок слов [Word order]		
	SVO	SOV	VSO
Гнев [Anger]	19	23	32
Отвращение [Disgust]	10	7	4
Страх [Fear]	7	19	20
Радость [Joy]	8	19	13
Грусть [Sadness]	15	16	13
Удивление [Surprise]	6	12	18
Всего [Total]	65	96	100

Таблица 8

**Количество примеров для каждой эмоции
в татарской обучающей выборке
в зависимости от порядка слов в предложении**
[Number of examples for each emotion in the Tatar training dataset
by the word order in the sentence]

Эмоция [Emotion]	Порядок слов [Word order]		
	SVO	SOV	VSO
Гнев [Anger]	21	24	5
Отвращение [Disgust]	15	29	11
Страх [Fear]	5	15	2
Радость [Joy]	27	47	16
Грусть [Sadness]	37	69	22
Удивление [Surprise]	21	23	5
Всего [Total]	126	207	61

В таблице 9 для русской тестовой выборки указано количество примеров для каждого порядка слов в зависимости от указанной эмоции. Для определения порядка слов была использована библиотека для обработки естественного языка SpaCy. Были выделены порядки SVO, SOV и VSO, при этом учитывались только примеры, где все три члена предложения (подлежащее, сказуемое и прямое дополнение) были выражены эксплицитно.

Таблица 9

**Количество примеров для каждой эмоции
в русской тестовой выборке
в зависимости от порядка слов в предложении
[Number of examples for each emotion in the Russian test dataset
by the word order in the sentence]**

Эмоция [Emotion]	Порядок слов [Word order]		
	SVO	SOV	VSO
Гнев [Anger]	86	30	19
Отвращение [Disgust]	36	13	18
Страх [Fear]	54	14	6
Радость [Joy]	66	6	22
Грусть [Sadness]	16	10	10
Удивление [Surprise]	14	12	6
Всего [Total]	272	85	81

Статистика порядка слов в указанных языках в целом совпадает с традиционно принятой классификацией базовых порядков. У немецкого и испанского языков количество примеров для порядков SVO и SOV находится на сравнительно равном уровне с небольшим преобладанием примеров с порядком SVO. В китайской обучающей выборке почти не наблюдается примеров с порядком SOV, а преобладающим порядком является SVO. Тестовая выборка на языке хинди имеет схожее количество примеров с порядками SOV и VSO, а также значительное количество примеров с порядком SVO. Наконец, татарская обучающая и русская тестовая выборки имеют ожидаемое преобладание примеров с порядком SOV и SVO соответственно.

6. Результаты и анализ

В данном разделе рассмотрим качество переноса модели, обученной на разных языках. Общее качество переноса языковых моделей на русский язык представлено в таблице 10, в таблице 11 для каждой модели представлено качество классификации отдельных эмоций.

Таблица 10

Общее качество классификации модели XLM-RoBERTa-base при ее дообучении на соответствующем исходном языке и последующем тестировании на русской выборке
[Overall classification performance of the XLM-RoBERTa-base model fine-tuned on each respective source language and evaluated on the Russian dataset]

Язык обучения [Language used for training]	Взвешенная F-мера [Weighted F-score]
Немецкий [German]	0,43
Испанский [Spanish]	0,63
Китайский [Chinese]	0,39
Хинди [Hindi]	0,55
Татарский [Tatar]	0,51

Таблица 11

Качество классификации отдельных эмоций в русских тестовых примерах в зависимости от исходного языка обучения модели
[Classification performance for individual emotions in the Russian test dataset by the source language used for model training]

Эмоция [Emotion]	Язык обучения [Language used for training]				
	Немецкий [German]	Испанский [Spanish]	Китайский [Chinese]	Хинди [Hindi]	Татарский [Tatar]
Гнев [Anger]	0,57	0,56	0,57	0,57	0,55
Отвращение [Disgust]	0,26	0,20	0,00	0,08	0,11
Страх [Fear]	0,43	0,81	0,00	0,71	0,27
Радость [Joy]	0,68	0,77	0,69	0,75	0,68
Грусть [Sadness]	0,09	0,69	0,39	0,39	0,68
Удивление [Surprise]	0,00	0,65	0,00	0,56	0,47

Наибольшее качество переноса показывают модели, дообученные на испанском языке и хинди. Скорее всего, модели, обученные на немецком и китайском языке, показывают более плохое качество в первую очередь из-за меньшего количества обучающих примеров для отдельных эмоций. Однако, несмотря на малое количество обучающих примеров для татарского языка, обученная на нем модель показала сравнительно высокое качество классификации, что, вероятнее всего, вызвано более близкой связью татарского языка с русским, а также схожестью их алфавитов.

Основное исследование состояло в проверке влияния порядка слов различных языков на качество переноса моделей, обученных на этих языках, на русские тексты. Для каждой перенесенной языковой модели были подсчитано ее качество классификации в зависимости от базового порядка слов. Полученные результаты приведены в таблице 12, на рисунке 1 представлен их график.

Таблица 12

**Общее качество классификации русских тестовых примеров
в зависимости от исходного языка обучения модели
и порядка слов в тексте**
[Overall classification performance on the Russian test dataset
by the source language and the word order in the text]

Язык обучения [Language used for training]	Порядок слов [Word order]		
	SVO	SOV	VSO
Немецкий [German]	0,55	0,41	0,43
Испанский [Spanish]	0,64	0,61	0,56
Китайский [Chinese]	0,34	0,28	0,32
Хинди [Hindi]	0,60	0,50	0,46
Татарский [Tatar]	0,41	0,47	0,38

Для каждого языка была оценена статистическая значимость порядка слов в предложении. В таблице 13 представлены значения *p-value*, показывающие значимость различия результатов классификации в зависимости от порядка слов в тестовых данных. В качестве метода использовался тест PERMANOVA [Anderson, 2001], который является непараметрическим пермутационным тестом оценки статистической значимости распределения в случае, когда имеет больше двух групп значений.

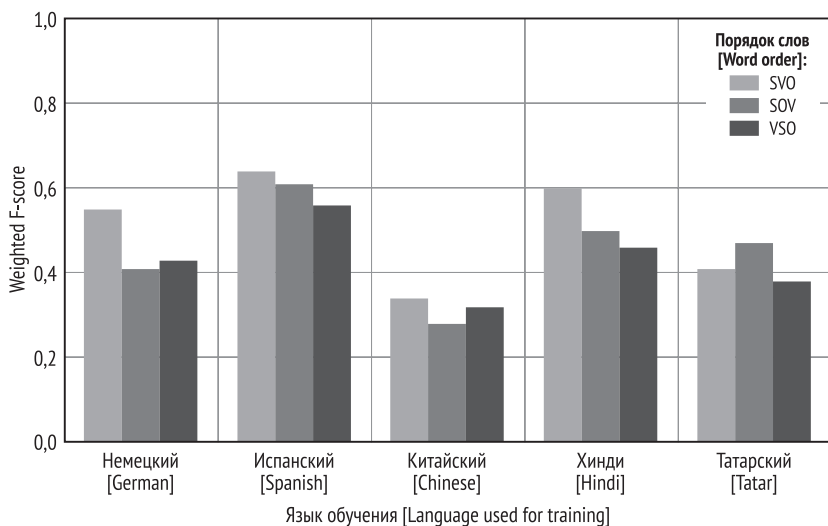


Рис. 1. График общего качества классификации русских тестовых примеров в зависимости от исходного языка обучения модели и порядка слов в тексте

Fig. 1. Graph showing the overall classification performance on the Russian test dataset by the source language and the word order in the text

Таблица 13

Результаты теста PERMANOVA для вычисления статистической значимости порядка слов при переносе обученной на соответствующем исходном языке модели на русский язык [PERMANOVA test results for assessing the statistical significance of word order when transferring a model trained on each respective source language to Russian]

Язык обучения [Language used for training]	p-value
Немецкий [German]	0,233
Испанский [Spanish]	0,444
Китайский [Chinese]	0,787
Хинди [Hindi]	0,108
Татарский [Tatar]	0,24

В качестве критического значения p -value здесь и во всех дальнейших подсчетах использовалось значение 0,05. Полученные в данном случае результаты указывают на отсутствие статистической значимости порядка слов в данных (p -value > 0,05 во всех случаях) для общей классификации эмоций. Модели, обученные на немецком, испанском и китайском языках, показывают наилучшее качество классификации для предложений, имеющих порядок SVO (см. рис. 1). В то же время у модели, обученной на татарском языке, можно увидеть увеличение качества при классификации предложений с порядком SOV. Из этого можно сделать предположение, что, хотя базовый порядок слов и не имеет достаточной статистической значимости при межъязыковом переносе моделей, в случае татарского языка базовый порядок слов все равно оказывает влияние на классифицирующую способность модели. Однако такой картины не наблюдается при обучении модели на языке хинди, тоже имеющим базовый порядок SOV. Вероятно, в данном случае на качество переноса также влияет близость языка-источника и языка-цели. Для более четкой картины необходимо провести дополнительные исследования с увеличенным объемом данных.

Для составления более четкой картины дополнительно было протестировано качество классификации для отдельных порядков слов в зависимости от эмоции, выраженной в предложении. В таблице 14 для каждого языка обучения представлено качество классификации примеров в зависимости от двух параметров: порядка слов и выражаемой эмоции. На основе этих данных на рисунке 2 представлены диаграммы для каждой отдельной эмоции.

Таблица 14

**Качество классификации отдельных эмоций
в русских тестовых примерах в зависимости
от исходного языка обучения модели и порядка слов в тексте**
[Classification performance for individual emotions in the Russian test
dataset by the source language and the word order in the text]

Эмоция [Emotion]	Язык обучения [Language used for training]	Порядок слов [Word order]		
		SVO	SOV	VSO
Гнев [Anger]	Немецкий [German]	0,75	0,97	0,73
	Испанский [Spanish]	0,69	0,70	0,88

Продолжение табл. 14

Эмоция [Emotion]	Язык обучения [Language used for training]	Порядок слов [Word order]		
		SVO	SOV	VSO
	Китайский [Chinese]	0,78	0,93	0,81
	Хинди [Hindi]	0,85	0,93	0,88
	Татарский [Tatar]	0,74	0,91	0,81
Отвращение [Disgust]	Немецкий [German]	0,78	0,63	0,62
	Испанский [Spanish]	0,24	0,70	0,20
	Китайский [Chinese]	0,00	0,00	0,00
	Хинди [Hindi]	0,11	0,00	0,00
	Татарский [Tatar]	0,00	0,00	0,00
Страх [Fear]	Немецкий [German]	0,62	0,44	0,50
	Испанский [Spanish]	0,94	0,73	1,00
	Китайский [Chinese]	0,00	0,00	0,00
	Хинди [Hindi]	0,77	0,60	0,80
	Татарский [Tatar]	0,31	0,25	0,50
Радость [Joy]	Немецкий [German]	0,90	0,80	0,90
	Испанский [Spanish]	0,94	1,00	0,90
	Китайский [Chinese]	0,86	0,80	0,71
	Хинди [Hindi]	0,94	1,00	0,84
	Татарский [Tatar]	0,71	1,00	0,62

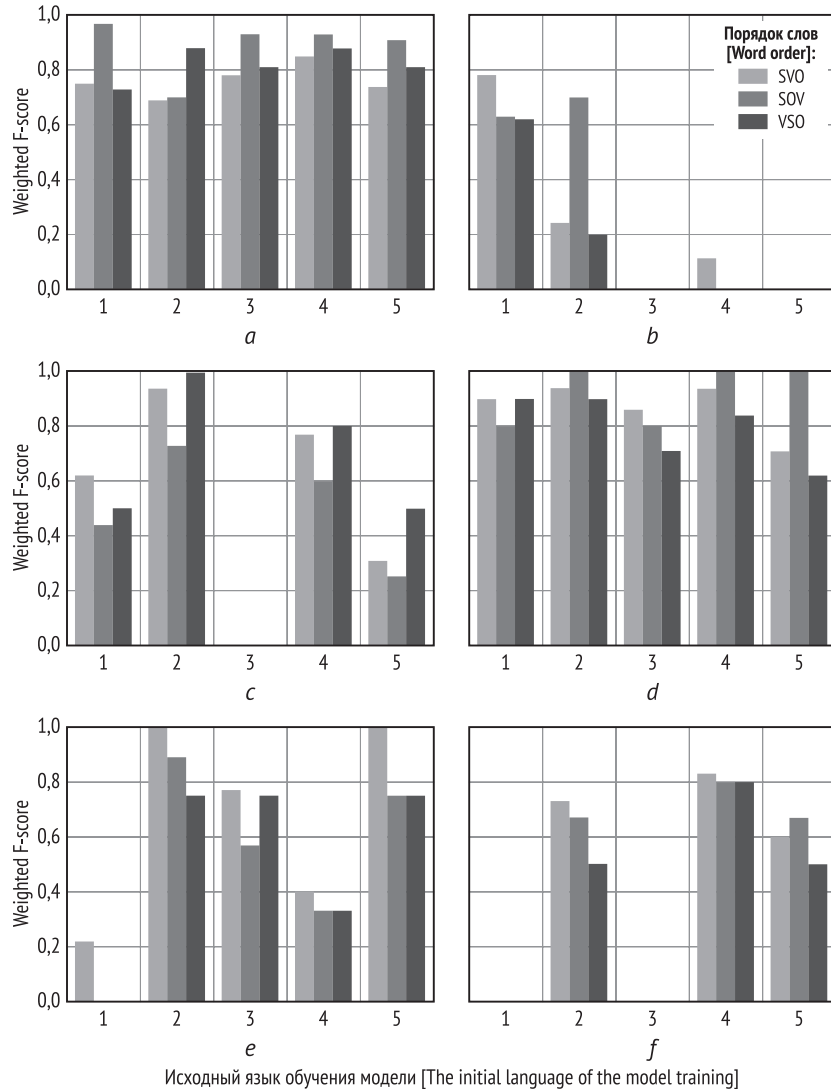
Эмоция [Emotion]	Язык обучения [Language used for training]	Порядок слов [Word order]		
		SVO	SOV	VSO
Грусть [Sadness]	Немецкий [German]	0,22	0,00	0,00
	Испанский [Spanish]	1,00	0,89	0,75
	Китайский [Chinese]	0,77	0,57	0,75
	Хинди [Hindi]	0,40	0,33	0,33
	Татарский [Tatar]	1,00	0,75	0,75
Удивление [Surprise]	Немецкий [German]	0,00	0,00	0,00
	Испанский [Spanish]	0,73	0,67	0,50
	Китайский [Chinese]	0,00	0,00	0,00
	Хинди [Hindi]	0,83	0,80	0,80
	Татарский [Tatar]	0,60	0,67	0,50

В целом, на основании полученных результатов сложно сделать однозначные выводы по причине малого количества примеров для некоторых эмоций, например, для эмоций отвращения и удивления. Однако рассмотрим основные факторы, влияющие на работу модели в данных условиях. На общей картине видно, что наилучшее качество предсказания наблюдается у эмоций гнева и радости, для которых имеется наибольшее количество примеров, а также эти эмоции обычно наиболее ярко выражены в тексте за счет лексики и пунктуации. Для других же эмоций картина более неоднозначная.

Для эмоции гнева хорошее качество классификации наблюдается у всех моделей, при этом более высокое качество наблюдается у предложений с порядком SOV. При более близком рассмотрении результатов одним из факторов является позиция окончания предложения, в которой при таком порядке оказывается глагол, часто несущий эмоциональную оценку.

Представляется, что на эту позицию модель обращает наибольшее внимание, ср. следующие предложения:

- 1) *Ублюдки бесят меня на протяжении всего дня!!!!;*
- 2) *Ты меня бесишь!!! Очень сильно!!!;*
- 3) *Всё, этот говнюк меня раззадорил. Меня еще никто так не бесил.*



Первые два предложения правильно классифицируются только двумя из пяти моделей, тогда как третье предсказывается правильно уже четырьмя моделями, за исключением испанской. Все предложения имеют один и тот же глагол *бесить*, 1-е и 3-е предложения имеют также оценочные слова *ублюдки* и *говнюк*, и при этом только 3-е предложение не имеет восклицательных знаков, часто свойственных для эмоции гнева, но все равно распознается лучше двух остальных за счет финальной позиции глагола.

Эмоция радости стабильно хорошо распознается всеми моделями, однако качество классификации у татарской модели заметно хуже других. В данном случае татарская модель предсказывает неправильную эмоцию гнева даже для таких сравнительно простых предложений, как *Надо же, они подняли настроение!* и *Я сделала это!*, с которыми справляются все остальные модели. К сожалению, однозначную причину этого установить сложно, и, скорее всего, это вызвано особенностью обучающих данных.

Для эмоций отвращения, страха и удивления высокое качество классификации наблюдается у языков, для которых в обучающей выборке было представлено большое количество примеров с соответствующей эмоцией. Так, качество классификации эмоции отвращения является наивысшим у немецкой модели, содержащей в обучающей выборке 294 примера с данной эмоцией, тогда как качество классификации эмоции страха выше всего у испанской модели, содержащей в обучающей выборке 295 примеров с данной эмоцией. В то же время на качество переноса здесь также влияет близость языков, например, испанская модель единственная правильно классифицировала предложение *Я стал виртуальным человеком, что меня и стало пугать*, в котором единственным показателем эмоции является глагол *пугать*, при том, что для модели на языке хинди имеется большее количество обучающих примеров для эмоции страха.

Рис. 2. Графики качества классификации отдельных эмоций в русских тестовых примерах в зависимости от исходного языка обучения модели и порядка слов в тексте:

a – гнев; *b* – отвращение; *c* – страх; *d* – радость; *e* – грусть; *f* – удивление
1 – немецкий; 2 – испанский; 3 – китайский; 4 – хинди; 5 – татарский

Fig. 2. Graphs showing the classification performance for individual emotions in the Russian test dataset by the source language and the word order in the text:

a – anger; *b* – disgust; *c* – fear; *d* – joy; *e* – sadness; *f* – surprise
1 – German; 2 – Spanish; 3 – Chinese; 4 – Hindi; 5 – Tatar

Порядок слов здесь опять же может оказывать влияние в зависимости от того, что находится на конце предложения, например, предложение *Самое страшное в жизни – это когда твоя семья тебя не понимает*(((испанская модель классифицирует как содержащее эмоцию «грусть» из-за наличия на конце грустных скобок (ср. предложение *Блин! могу сидеть только в твиттере, а так хочется с тобой поговорить*(((или *чтобы ты мне написал!!!*, которое содержит на конце три восклицательных знака и которое та же испанская модель классифицирует с эмоцией радости).

Наконец, для эмоции грусти наблюдается несколько иная картина. Хотя все языки имеют достаточное большое количество обучающих примеров для данной эмоции, некоторые модели при переносе показывают крайне плохие результаты классификации, особенно немецкая модель. Это обусловлено в первую очередь тем, что основным способ выражения грусти в русских тестовых предложениях – это несколько идущих подряд круглых скобок: (((. Это приводит к тому, что лучше всего такие примеры классифицируют модели, обученные на более близких языках или языках, имеющих схожую традицию передачи эмоций в тексте. Что интересно, среди данных языков только татарский имеет обучающие примеры, также содержащие несколько идущих подряд круглых скобок, но при этом испанская модель все равно способна их распознавать даже с более высоким результатом. Основным средством выражения эмоции грусти в испанской выборке является такой смайл: :(. Он встречается в большинстве примеров с эмоцией грусти (в отличие от других языков, где не наблюдается специфических смайлов или эмодзи). Высокое качество переноса для испанского языка в данном случае указывает на высокую способность модели к генерализации.

Полученные результаты, таким образом, указывают на отсутствие значимого влияния базового порядка слов языка-источника на качество переноса обученных моделей, но при этом порядок слов все еще оказывает влияние на классификацию: в первую очередь на это влияют слова и эмоциональные маркеры, стоящие в конце предложения. Тем не менее, основное влияние на качество переноса языковых моделей по-прежнему оказывает количество и качество обучающих данных моделей, а также близость языка-источника и языка-цели.

8. Заключение

В данной работе был рассмотрен перенос языковых моделей, предварительно обученных на одном языке, на другой язык в рамках задачи автоматической классификации эмоции в текстах на основе данных, представленных на соревновании SemEval-2025.

В рамках практического исследования был проведен анализ влияния базового порядка слов на качество классификации языковых моделей, обученных на немецком, испанском, китайском, хинди и татарском языках и затем протестированных на данных русского языка. Для полученных результатов также была подсчитана статистическая значимость.

На основе анализа работы языковых моделей и имеющихся данных были описаны основные факторы, влияющие на качество переноса рассмотренных моделей, в том числе и влияние порядка слов. Одним из основных выводов является отсутствие значимого влияния базового порядка слов в языке-источнике, но наблюдается влияние слов и/или знаков пунктуации, стоящих в финальной позиции предложения.

Несмотря на это, имеющихся эмоциональных текстов недостаточно для проведения полноценного всестороннего исследования влияния базового порядка слов на качество переноса языковых моделей, и дальнейшим направлением работы представляется сбор большего объема данных, а также потенциально более систематизированный подход к их отбору в зависимости от порядка слов и наличия в них тех или иных эмоциональных маркеров.

Библиографический список / References

Anderson, 2001 – Anderson M.J. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*. 2001. Vol. 26. No. 1. Pp. 32–46.

Conneau et al., 2020 – Conneau A., Khandelwal K., Goyal N. et al. Unsupervised cross-lingual representation learning at scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. D. Jurafsky, J. Chai, N. Schlueter, J. Tetreault (eds.). Association for Computational Linguistics, 2020. Pp. 8440–8451.

Cowie et al., 2001 – Cowie R., Douglas-Cowie E., Tsapatsoulis N. et al. Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*. 2001. Vol. 18. No. 1. Pp. 32–80.

Deshpande et al., 2022 – Deshpande A., Talukdar P., Narasimhan K. When is BERT multilingual? Isolating crucial ingredients for cross-lingual transfer. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. M.C. de Marneffe, I.V. Meza Ruiz (eds.). Seattle, USA, 2022. Pp. 3610–3623.

Devlin et al., 2019 – Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. J. Burstein, C. Doran, T. Solorio (eds.). Minneapolis, Minnesota, 2019. Pp. 4171–4186.

Dryer, Haspelmath, 2013 – Dryer M.S., Haspelmath M. The world atlas of language structures online. *Zenodo*. Leipzig, 2024. DOI: 10.5281/zenodo.13950591.

Hansen et al., 2022 – Hansen L., Zhang Y.P., Wolf D. et al. A generalizable speech emotion recognition model reveals depression and remission. *Acta Psychiatrica Scandinavica*. 2022. Vol. 145. No. 2. Pp. 186–199.

Jackson et al., 2019 – Jackson J.K., Watts J., Henry T.R. et al. Emotion semantics show both cultural variation and universal structure. *Science*. 2019. Vol. 366. No. 6472. Pp. 1517–1522.

Kovacevic et al., 2022 – Kovacevic L., Bradic V., Melo G. et al. EZGlot. 2022. URL: <https://ezglot.com> (date accessed: 18.08.2025).

Lauscher et al., 2020 – Lauscher A., Ravishankar V., Vulić I., Glavaš G. From zero to hero: On the limitations of zero-shot language transfer with multilingual transformers. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. B. Webber, T. Cohn, Y. He, Y. Liu (eds.). Association for Computational Linguistics, 2020. Pp. 4483–4499.

Lefter, Jonker, 2017 – Lefter I., Jonker C.M. Aggression recognition using overlapping speech. *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. San Antonio, TX, 2017. Pp. 299–304.

Lin et al., 2019 – Lin Y.H., Chen C.Y., Lee J. et al. Choosing transfer languages for cross-lingual learning. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. A. Korhonen, D. Traum, L. Márquez (eds.). Florence, Italy, 2019. Pp. 3125–3135.

Littell et al., 2017 – Littell P., Mortensen D.R., Lin K. et al. URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. M. Lapata, P. Blunsom, A. Koller (eds.). Valencia, Spain, 2017. Pp. 8–14.

Liu et al., 2019 – Liu Y., Ott M., Goyal N. et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv.org*. 2020. DOI: 10.48550/arXiv.1907.11692.

Liu et al., 2020 – Liu C.L., Hsu T.Y., Chuang Y.S., Lee H.Y. A study of cross-lingual ability and language-specific information in multilingual BERT. *arXiv.org*. 2020. DOI: 10.48550/arXiv.2004.09205.

Mitsios et al., 2024 – Mitsios M., Vamvoukakis G., Maniati G. et al. Improved text emotion prediction using combined valence and arousal ordinal classification. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*. K. Duh, H. Gomez, S. Bethard (eds.). Mexico City, Mexico, 2024. Pp. 808–813.

Muhammad et al., 2025a – Muhammad S.H., Ousidhoum N., Abdulmumin I. et al. BRIGHTER: BRIDging the Gap in Human-Annotated Textual Emotion Recognition datasets for 28 languages. *arXiv.org*. 2025. DOI: 10.48550/arXiv.2502.11926.

Muhammad et al., 2025b – Muhammad S.H., Ousidhoum N., Abdulmumin I. et al. SemEval-2025 Task 11: Bridging the gap in text-based emotion detection. *arXiv.org*. 2025. DOI: 10.48550/arXiv.2503.07269.

Patil et al., 2022 – Patil V., Talukdar P., Sarawagi S. Overlap-based vocabulary generation improves cross-lingual transfer among related languages. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. S. Muresan, P. Nakov, A. Villavicencio (eds.). Dublin, Ireland, 2022. Pp. 219–233.

Philippy et al., 2023 – Philippy F., Guo S., Haddadan S. Towards a common understanding of contributing factors for cross-lingual transfer in multilingual language models: A review. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. A. Rogers, J. Boyd-Graber, N. Okazaki (eds.). Toronto, Canada, 2023. Pp. 5877–5891.

Pikuliak et al., 2021 – Pikuliak M., Šimko M., Bieliková M. Cross-lingual learning for text processing: A survey. *Expert Systems with Applications*. 2021. Vol. 165. No. 113765. DOI: 10.1016/j.eswa.2020.113765.

Radford et al., 2018 – Radford A., Narasimhan K., Salimans T., Sutskever I. Improving language understanding by generative pre-training. 2018. URL: <https://gwern.net/doc/www/s3-us-west-2.amazonaws.com/d73fdc5ffa8627bce44dcda2fc012da638ffb158.pdf> (date accessed: 18.08.2025).

Vaswani et al., 2017 – Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need. *arXiv.org*. 2017. DOI: 10.48550/arXiv.1706.03762.

Wichmann et al., 2010 – Wichmann S., Holman E.W., Bakker D., Brown C.H. Evaluating linguistic distance measures. *Physica A: Statistical Mechanics and its Applications*. 2010. Vol. 389. No. 17. Pp. 3632–3639.

Wu et al., 2023 – Wu Z., Tamkin A., Papadimitriou I. Oolong: Investigating what makes transfer learning hard with controlled studies. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. H. Bouamor, J. Pino, K. Bali (eds.). Singapore, 2023. Pp. 3280–3289.

Статья поступила в редакцию 10.06.2025

The article was received 10.06.2025

Сведения об авторах / About the authors

Лемаев Владислав Игоревич – аспирант кафедры теоретической и прикладной лингвистики филологического факультета, Московский государственный университет имени М.В. Ломоносова

Vladislav I. Lemaev – PhD student at the Department of Theoretical and Applied Linguistics of the Faculty of Philology, Lomonosov Moscow State University, Moscow, Russian Federation

E-mail: vladzhkv98@mail.ru

Наталья Валентиновна Лукашевич – доктор технических наук, кандидат физико-математических наук; профессор кафедры теоретической и прикладной лингвистики филологического факультета, Московский государственный университет имени М.В. Ломоносова

Natalia V. Loukashevitch – Dr. Hab. in Technical Sciences, PhD in Mathematics; Professor at the Department of Theoretical and Applied Linguistics of the Faculty of Philology, Lomonosov Moscow State University, Moscow, Russian Federation

E-mail: louk_nat@mail.ru