



УДК 81-139 + 612.821

EDN OKRCHB

<https://www.doi.org/10.33910/2687-0223-2023-5-1-37-45>

Применение психофизиологического инструментария для оценки согласованности в разметке тональности текстов с переключением кода на основе романа Сергея Минаева

О. Д. Рябова^{✉1}

¹ Новосибирский национальный исследовательский государственный университет, 630090, Россия, г. Новосибирск, ул. Пирогова, д. 1

Сведения об авторе

Рябова Ольга Денисовна,
ORCID: [0000-0003-1806-3555](https://orcid.org/0000-0003-1806-3555),
e-mail: o.ryabova@ngsu.ru

Для цитирования:

Рябова, О. Д.
(2023) Применение психофизиологического инструментария для оценки согласованности в разметке тональности текстов с переключением кода на основе романа Сергея Минаева. *Комплексные исследования детства*, т. 5, № 1, с. 37–45.
<https://doi.org/10.33910/2687-0223-2023-5-1-37-45> EDN OKRCHB

Получена 15 января 2023; прошла рецензирование 22 января 2023; принята 23 января 2023.

Финансирование: Исследование не имело финансовой поддержки.

Права: © О. Д. Рябова (2023). Опубликовано Российским государственным педагогическим университетом им. А. И. Герцена. Открытый доступ на условиях лицензии [CC BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

Аннотация. Переключение кода — это явление, при котором в одном и том же сообщении используются два или более языка. В настоящее время в социальных сетях, дискурсе IT-специалистов и билингвов довольно часто можно встретить сообщения на смешанных языках. Этот языковой феномен представляет собой проблему для ряда задач обработки естественного языка, включая анализ настроений.

В данной статье изучается согласованность разметки тональности текстов с переключением на латиницу экспертами с помощью инструментов, которые показали высокую эффективность в междисциплинарных исследованиях. К числу таких методов относится 2B-PLS, обеспечивающий дополнительные возможности для анализа современных экспериментов в нейролингвистике, психофизиологии и других областях наук.

Для примера был выбран роман Сергея Минаева «The Телки. Повесть о ненастоящей любви», который был издан в 2008 году. Были выбраны 100 предложений, как из речи автора, так и из диалогов персонажей, содержащие в себе слова, написанные на латинице. Из предложений был собран и размечен датасет в формате CSV для дальнейшего построения модели.

Параметрами для 2B-PLS анализа послужили результаты экспертной оценки тональности выбранных предложений: количество сущностей на латинице и всего во фразе соответственно, согласованность в оценках экспертов позитивной, негативной и нейтральной тональностей, признаки категорий сущностей на латинице Локация, Персона, Время/Дата, Бренд, Организация, Модель, незначащая сущность соответственно.

2B-PLS анализ показал возможность анализа согласованности оценок тональности фразы экспертами в зависимости от знаний, извлеченных из предложений: именованных сущностей и другой статистики. На согласованность оценок экспертов влияет не только сама категория сущностей, но и тональность фразы, а также общее количество сущностей и сущностей на латинице во фразе.

Полученные результаты соотносятся с теоретическими исследованиями.

Ключевые слова: 2B-PLS, переключение кода, именованные сущности, анализ тональности, Сергей Минаев, извлечение информации из текстов, извлечение знаний из текстов

The use of psychophysiological tools for assessing consistency in the markup of the tonality of texts with code-switching based on a novel by Sergey Minaev

O. D. Ryabova^{✉1}

¹ Novosibirsk National Research State University, 1 Pirogova Str., Novosibirsk 630090, Russia

Author

Olga D. Ryabova,
ORCID: [0000-0003-1806-3555](https://orcid.org/0000-0003-1806-3555),
e-mail: o.ryabova@ng.nsu.ru

For citation:

Ryabova, O. D.
(2023) The use
of psychophysiological tools for
assessing consistency in the markup
of the tonality of texts with code-
switching based on a novel
by Sergey Minaev. *Comprehensive
Child Studies*, vol. 5, no. 1, pp. 37–45.
[https://doi.org/10.33910/2687-
0223-2023-5-1-37-45](https://doi.org/10.33910/2687-0223-2023-5-1-37-45)
EDN OKRCHB

Received 15 January 2022;
reviewed 22 January 2023;
accepted 23 January 2023.

Funding: The study did not receive
any external funding.

Copyright: © O. D. Ryabova (2023).
Published by Herzen State
Pedagogical University of Russia.
Open access under [CC BY-NC
License 4.0](https://creativecommons.org/licenses/by-nc/4.0/).

Abstract. Code-switching is a phenomenon in which two or more languages occur in the same message. Messages containing mixed languages are quite common in social networks, as well as the discourse of IT specialists and bilinguals. Code-switching presents a challenge for sentiment analysis and other natural language processing tasks.

This article explores the possibility of identifying complex cases in the markup of texts with code-switching using a PLS analysis. Sergey Minaev's 2008 novel *The Chicks. A Tale of Unreal Love* was chosen for the analysis. 100 sentences containing words written in Latin were selected, both from the author's speech and from the dialogues of the characters. A dataset was collected and marked up in the CSV format for further model construction.

The parameters for the analysis were the results of an expert assessment of the sentiments of the selected sentences: the number of entities in Latin and the total number of entities in the phrase; consistency in expert assessments of positive, negative and neutral sentiments; features of categories of entities in Latin (Location, Person, Time/Date, Brand, Organization, Model); an insignificant entity.

The 2B-PLS analysis showed that it is possible to analyze the consistency of sentiment markup of a phrase by experts depending on the knowledge extracted from the sentences—i.e., named entities and other statistics. The consistency of expert assessments is influenced not only by the category of entities, but also by the sentiment of the phrase, as well as the total number of entities and entities in Latin in the phrase.

The results obtained correlate with theoretical studies.

Keywords: 2B-PLS, code-switching, named entities, sentiment analysis, Sergey Minaev, information extraction, text mining

Введение

Человеку присущи выбор и построение языковых регистров в зависимости от коммуникативной ситуации, а также способность к созданию и изучению новых методов коммуникации. Переключение кода в общении между людьми используется для полного выражения своих мыслей и чувств или приспособления к новому коллективу. В общении между человеком и машиной, таком как взаимодействие с виртуальным помощником Alexa или Алиса, переключение между языками представляет сложную проблему. Если мы хотим, чтобы машины участвовали в подобных взаимодействиях, необходимо уметь понимать, что происходит в языковых регистрах и с какими трудностями могут встретиться языковые системы.

В литературе существует больше публикаций о переключении кодов между английским и каким-либо другим языком среди английских билингвов, нежели среди русско-английских

билингвов. Для русского как гостевого языка (прим. неродной язык для носителя) написано больше работ про переключение, например, с казахского на русский (Akishev 2020) или с иврита на русский (Naiditch 2000). Это легко объясняется влиянием Советского Союза на языковую политику республик. После распада СССР многие русскоговорящие люди иммигрировали в США, поэтому сегодня поколение их детей и внуков являются теми носителями языка наследия, которые представляют исследовательский интерес в данной статье.

Для русского как матричного языка (прим. родной язык носителя) и английского как гостевого в последние годы проведен ряд исследований. В одном из исследований (Burdygina 2021) с использованием методологии матричной языковой фреймовой модели Майерса-Скоттона (1993) проанализированы языковые единицы московских IT-специалистов, которые представлены кодовыми переключениями с точки зрения их синтаксических, морфологических

особенностей и прагматических функций. Цель другого исследования (Kuleshova 2020) состояла в том, чтобы подсчитать и проанализировать случаи переключения кода между русским и английским языками среди американских школьников с русскоговорящими родителями и доказать или отвергнуть гипотезу о том, что в диалоге, посвященном семейному празднику, проявится меньше случаев переключения кода по сравнению с диалогом, посвященным занятиям в университете. Темы для диалога определили заранее, при этом никакой подготовки не требовалось. Речь должна была быть спонтанной и естественной. Полученные результаты подтвердили гипотезу о том, что случаи переключения кода в первом диалоге происходили менее часто, то есть студенты в основном использовали только один язык — русский, не очень часто переключаясь на английский. Второй диалог выявил больше случаев переключения кода, как и предполагалось изначально.

В обработке естественно-языковых (ЕЯ) текстов можно выделить два главных направления: извлечение информации из текстов (Information Extraction, IE) и извлечение знаний из текстов (Text Mining) (Najdenova, Nevzorova 2008). Определяя задачу извлечения информации, исследователи заметили, что важно распознавать информационные единицы, такие как имена людей, названия организаций и местоположений, а также числовые выражения, включая время, дату, деньги и процентные выражения (Nadeau, Sekine 2007). Идентификация ссылок на эти объекты в тексте была признана одной из важных подзадач IE и получила название «Распознавание именованных сущностей» (Named Entity Recognition, NER). Задача их выделения состоит в присвоении каждому токenu BIO-разметки: «В» — начало сущности, «И» — продолжение сущности, «О» — не относится к сущности. Для задачи NER в текстах с переключением кода были проведены исследования на материале постов из социальной сети для испанского-английского, арабского-английского (Aguilar et al. 2018) и хинди-английского (Singh et al. 2018). К задачам извлечения знаний относится анализ настроений (Sentiment Analysis) — область обработки естественного языка, компьютерной лингвистики и текстовой аналитики, имеющая целью выявление и классификацию субъективных мнений в исходных данных (например, в документе или предложении) (Luo et al. 2013).

В последние годы для текстов с переключением кода в вышеупомянутой задаче был проведен ряд исследований. Так, Б. Г. Патра, Д. Дас,

А. Дас (Patra et al. 2018) исследовали тональность для двуязычных пар на хинди-английском и бенгальском-английском и показали, что наиболее эффективная система анализа использовала признаки n-граммы на уровне слов и символов с классификатором, основанном на методе опорных векторов (Support Vector Machine, SVM).

Из проведенного обзора можно заключить, что для русского и английского переключения кода наблюдается недостаток исследований. Это представляет собой проблему: ряд авторов обнаружили корреляцию между использованием языка и тональностью, показав, что игнорирование одного языка в пользу другого или полное игнорирование языков с переключением кода может привести к неправильным выводам о настроениях пользователей (Agarwal et al. 2017). Таким образом, исследование переключения с русского (матричный язык) на английский (гостевой) в дискурсе как книжных персонажей, так и в дискурсе IT-специалистов или русских эмигрантов является актуальным дополнением к написанным работам.

Материалы и методы

Для анализа использовались 100 предложений из романа Сергея Минаева «The Тёлки. Повесть о ненастоящей любви». Произведение было выбрано в связи с вышедшей в 2022 году экранизацией «The Телки» — российским драматическим мини-сериалом по мотивам романа. Предложения были автоматически извлечены из книги в формате PDF с помощью языка программирования Python и библиотеки PyPDF2.

Была получена таблица в формате CSV со следующими колонками:

- 1) идентификатор предложения;
- 2) предложение;
- 3) сентимент предложения: позитивный/негативный/нейтральный для 9 экспертов;
- 4) номер предложения в тексте;
- 5) количество сущностей на латинице;
- 6) количество сущностей в целом;
- 7) количество символов на латинице;
- 8) количество букв и цифр в предложении;
- 9) количество знаков препинания и пробельных символов;
- 10) суммарное количество символов в предложении;
- 11) именованные сущности: «Локация», «Персона», «Время», «Бренд», «Организация», «Модель», «Незначащая сущность».

Необходимые расчеты производились при помощи скриптов, написанных на языке Python,

регулярных выражений, пакетов punkt и words библиотеки Natural Language Toolkit (nltk). Подробно опишем процедуру извлечения и разметки предложений.

После импортирования всех необходимых библиотек файл романа в формате PDF считывается с помощью функции PdfFileReader библиотеки PyPDF2. Эта функция позволяет получить количество страниц в документе и содержание каждой страницы в строковом формате. Для более удобного автоматического анализа текста PdfFileReader обрезает верхние заголовки и номера страниц. Затем мы проходим циклом по всем страницам в тексте и делим его на предложения с помощью библиотеки nltk и запоминаем его номер. Каждое предложение затем с помощью регулярных выражений очищается от типографских ошибок (лишних пробелов, пробелов после знаков препинания, повторяющихся знаков препинания) и токенизируется для подсчета сущностей и последующей разметки. Создается список из предложений, содержащих в себе символы на латинице с помощью регулярного выражения $r'[A-Za-z]+'$. Исключаются названия глав, написанные в верхнем регистре, и незначащие для анализа знаки препинания.

Затем предложения были размечены вручную экспертом в сфере компьютерной лингвистики и основам машинного обучения. Были выделены именованные сущности «Локация», «Персона», «Время», «Бренд», «Организация», «Модель», «Незначащая сущность» в соответствии со следующими правилами:

- 1) все помещения и места, в которых человек может физически находиться (бассейн, ресторан, лестница и т. п.), а также их названия, например «метро Аэропорт», отмечаются как сущность «Локация»;
- 2) все упоминания людей (имена, личные местоимения на русском и английском языках) отмечаются как сущность «Персона»;
- 3) все упоминания временных периодов и дат, например «четыре раза в неделю», «через две минуты», «пятого марта» и т. п., отмечаются как сущность «Время»;
- 4) все бренды одежды или аксессуаров, такие как Patrick Hellmann или Tiffany, отмечаются как сущность «Бренд»;
- 5) все компании с более широким ассортиментом и производством, такие как Walmart и ИКЕА, или организации, такие как новостные каналы, банки, журналы, другие магазины, отмечаются как сущность «Организация»;

- 6) было принято решение делить сущности на «Бренд» и «Организацию», а затем связать сущность «Бренд» и «Модель» для таких случаев, как «Hammer H2» или «Nokia 8800». Разметка в этом случае будет выглядеть как «B-brand, B-model».

Тональности предложений были получены с помощью оценки 9 экспертов, изучавших курс «Методы и алгоритмы компьютерной лингвистики». Исследование проводилось в облачной среде «Google Таблицы». Эксперты заполняли поле с фамилией, именем и отчеством, а затем следовали инструкции: «...оцените высказывания в каждой из 100 строк по признаку позитивное — нейтральное — негативное суждение (поставьте 1 напротив предложения в нужном столбце)». Таким образом, были получены оценки тональности предложений для блока предикторов 2B-PLS-модели.

Методы анализа

В последние годы при изучении имплицитных процессов различной природы интенсивно используется такой инструмент многомерного анализа, как 2B-PLS (Two-Block Projection to Latent Structure или Partial Least Squares). С его помощью эффективно определяются глубинные «латентные структуры» (независимые механизмы), которые едины и для блока предикторов (данные для модели), и для блока откликов (вопросы модели) с условием максимума шаблона ковариации между новыми матрицами счетов предикторов и откликов (Rohlf, Corty 2000; Rännar et al. 1994).

В модели 2B-PLS блок № 1 (матрица предикторов) после центрирования, масштабирования и поворота дает ответы на вопросы-отклики из блока № 2. А поиск угла для поворота предикторов (обучение модели) происходит с помощью матрицы откликов (блок № 2) — в этом физический смысл откликов-вопросов. Параметры трансформации от исходных матриц к новым матрицам счетов находятся в матрицах нагрузок предикторов и откликов. Матрицы нагрузок фактически содержат коэффициенты корреляции между исходными матрицами и найденными латентными структурами.

В междисциплинарных экспериментах 2B-PLS-анализ хорошо себя зарекомендовал в исследованиях в области психофизиологии (Krivoshechekov et al. 2022; Nikolaeva et al. 2022), нейронаук (Savostyanov et al. 2022), биологии (Kovaleva et al. 2019), психологии (Vergunov 2022), генетики (Polunin et al. 2019).

Модель формировалась с помощью программного пакета JACOBI 4 (Polunin et al. 2019).

Результаты

В 2B-PLS-модель вошли 5 переменных в блок № 1 и 6 признаков в блок № 2 (табл. 1), что дает 5 латентных структур в модели.

Таким образом, предикторы 2B-PLS-модели (блок № 1) должны дать ответы на вопросы о сущностях на латинице и кириллице в текстах, а также согласованности оценок экспертов. Обучение модели (выбор угла разворота блока данных) происходит за счет откликов (блок № 2), которые включают именованные сущности.

Как следует из графика осыпи латентных структур, сформированных моделью (рис. 1), все пять латентных структур превышают порог 5% описания дисперсии. Из них только структура 4 не показала связи с несогласованностью оценок экспертов. При этом структуры 1 и 3 связаны с несогласованностью в оценках негативной тональности, а структуры 2 и 5 — с несогласованностью в оценках нейтральной и позитивной тональностей.

Табл. 1. Блоки переменных для 2B-PLS-модели

Переменные	Блок
&n_lat_ent, &n_all_ent: количество сущностей на латинице и всего во фразе соответственно	№ 1
&Pos, &Neg, &Neut: согласованность в оценках экспертов позитивной, негативной и нейтральной тональностей соответственно	№ 1
_s_Loc, _s_Per, _s_Time, _s_Brand, _s_Org, _s_Model, _s_O: признаки категорий сущностей на латинице: Локация, Персона, Время/Дата, Бренд, Организация, Модель, незначащая сущность соответственно	№ 2

Примечание: переменные блока 1 предварительно нормировались на размах.

Table 1. Blocks of variables for 2B-PLS-model

Variables	Block
&n_lat_ent, &n_all_ent: the number of entities in Latin and total number of entities in the phrase, respectively	No. 1
&Pos, &Neg, &Neut: consistency in expert assessments of positive, negative and neutral tones, respectively	No. 1
_s_Loc, _s_Per, _s_Time, _s_Brand, _s_Org, _s_Model, _s_O: features of entity categories in Latin: Location, Person, Time/Date, Brand, Organization, Model, an insignificant entity, respectively	No. 2

Note: the variables of Block 1 were previously normalized to the range.

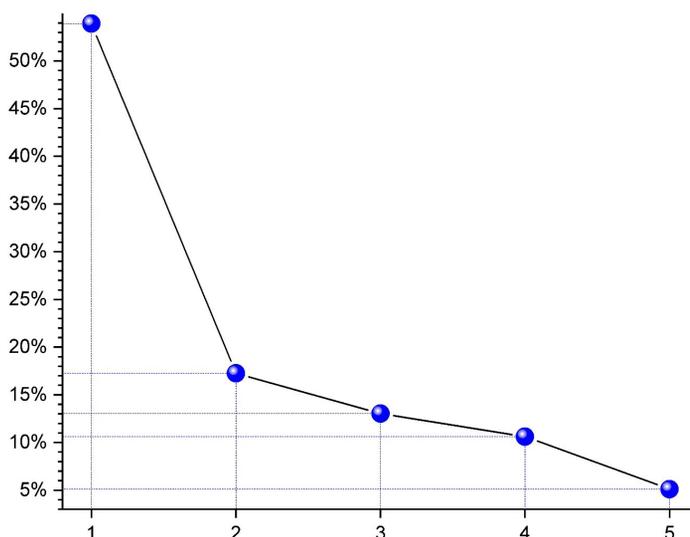


Рис. 1. График осыпи для латентных структур 2B-PLS-модели. Горизонтальная ось — номера латентных структур, вертикальная ось — описываемая доля общей наблюдаемой дисперсии

Fig. 1. Scree plot for latent structures of the 2B-PLS model. The horizontal axis is the numbers of latent structures, the vertical axis is the described fraction of the total observed variance

Как следует из рисунка 2, для латентных структур 1 и 3 получилось два кластера с центрами в признаке *_s_Org* и в признаке *_s_Per*.

Первый кластер показывает, что для категории сущностей «Организация» характерно появление во фразах с большим общим количеством сущностей.

Второй кластер показывает, что для категорий сущностей «Локация», «Персона» и «Бренд» характерна менее согласованная оценка экспертами фраз для случая негативной тональности.

Отметим, что на рисунке 2, хорошо заметна косоугольная система координат с осями «общее количество сущностей во фразе» и «согласованность оценок экспертов по негативной тональности». Однако из-за наличия угла наклона между осями (их коррелированности) сами эти оси не могут стать латентными структурами — все латентные структуры обязательно должны быть независимыми друг от друга (ортогональными). Но фактически наша переменная «общее количество сущностей во фразе» полностью определяет латентную структуру 1, а наша переменная «согласованность оценок экспертов

по негативной тональности» — полностью определяет латентную структуру 2 (у данных переменных самая тесная корреляция с осями соответствующих латентных структур).

Как следует из рисунка 3, для латентных структур 2 и 5 получилось три перекрывающихся кластера с центрами в признаках *_s_Per*, *_s_Model* и *_s_Loc*. Наша переменная «согласованность оценок экспертов по нейтральной тональности» полностью определяет латентную структуру 1, а наша переменная «согласованность оценок экспертов по позитивной тональности» — полностью определяет латентную структуру 2 (у данных переменных самая тесная корреляция с осями соответствующих латентных структур).

Первый кластер показывает, что для категории сущностей Персона характерна более согласованная оценка экспертами позитивной или нейтральной тональности во фразах с большим общим количеством сущностей.

Второй кластер показывает, что для категории незначащих сущностей, или «Модель», характерна более согласованная оценка экспертами фраз нейтральной тональности, а для фраз с позитивной тональностью — менее согласованная.

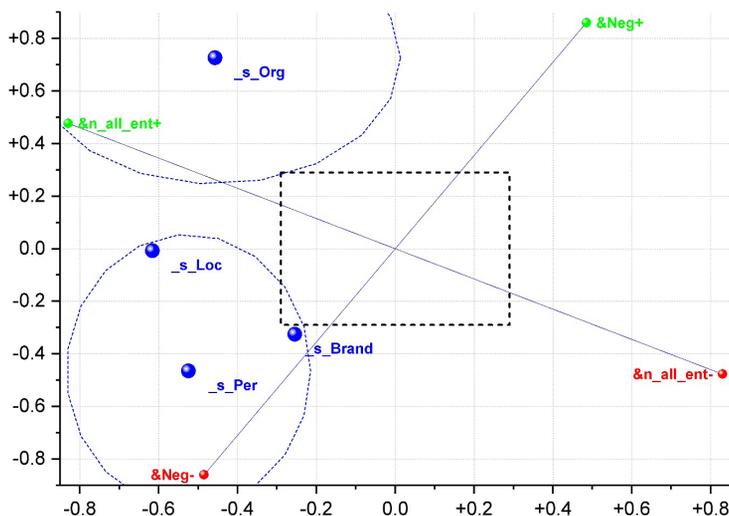


Рис. 2. Визуализация статистически значимых нагрузок для переменных (латентные структуры № 1 и № 3). Отметки зеленого цвета (название завершается знаком «+») показывают более высокие значения переменных, отметки красного цвета (название завершается знаком «-») показывают более низкие значения переменных; горизонтальная ось — коэффициенты корреляции переменных с латентной структурой № 1, вертикальная ось — с латентной структурой № 3; синим пунктиром даны области/кластеры переменных; вне области, ограниченной черным пунктиром, находятся переменные со статистически значимыми нагрузками; условные обозначения соответствуют таблице 1

Fig. 2. Visualization of statistically significant loadings for variables (latent structures No. 1 and No. 3). The green marks (the name ends with “+”) show higher values of variables, the red marks (the name ends with “-”) show lower values of variables; the horizontal axis is the correlation coefficients of variables with latent structure No. 1, the vertical axis, with latent structure No. 3; the blue dotted line shows the areas/clusters of variables; outside the area bounded by the black dotted line, there are variables with statistically significant loads; the symbols correspond to Table 1

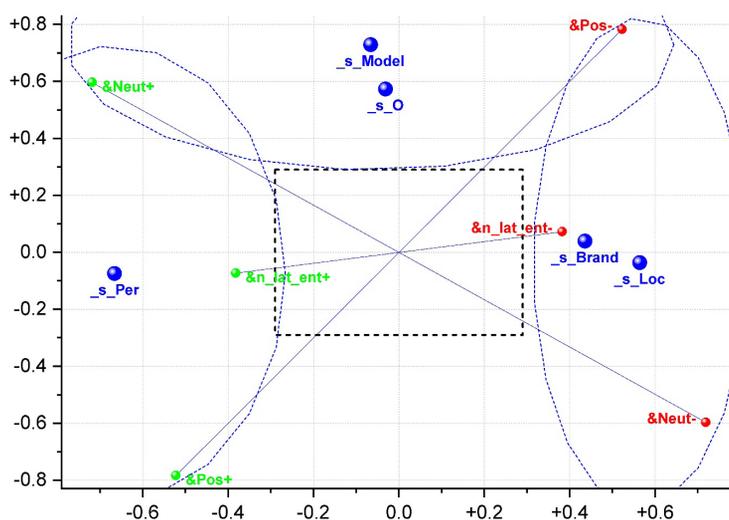


Рис. 3. Визуализация статистически значимых нагрузок для переменных (латентные структуры № 2 и № 5). Отметки зеленого цвета (название завершается знаком «+») показывают более высокие значения переменных, отметки красного цвета (название завершается знаком «-») показывают более низкие значения переменных; горизонтальная ось — коэффициенты корреляции переменных с латентной структурой № 2, вертикальная ось — с латентной структурой № 5; синим пунктиром даны области/кластеры переменных; вне области, ограниченной черным пунктиром, находятся переменные со статистически значимыми нагрузками; условные обозначения соответствуют таблице 1

Fig. 3. Visualization of statistically significant loadings for variables (latent structures No. 2 and No. 5). The green marks (the name ends with “+”) show higher values of variables, the red marks (the name ends with “-”) show lower values of variables; the horizontal axis is the correlation coefficients of variables with latent structure No. 2, the vertical axis, with latent structure No. 5; the blue dotted line shows the areas/clusters of variables; outside the area bounded by the black dotted line, there are variables with statistically significant loads; the symbols correspond to Table 1

Третий кластер показывает, что для категорий сущностей «Локация» и «Бренд» характерна менее согласованная оценка экспертами фраз нейтральной и позитивной тональности для случая с меньшим количеством сущностей на латинице.

Таким образом, на согласованность оценки экспертов влияет не только сама категория сущностей, но и тональность фразы, а также общее количество сущностей и сущностей на латинице во фразе.

Можно сделать вывод, что уточнение эксплицитных правил для негативной тональности сущностей «Локация», «Персона» и «Бренд», а также правил для нейтральной и позитивной тональности в случае малого количества сущностей на латинице позволит улучшить обучение нейросетей или разработать более эффективные алгоритмы машинного перевода.

Обсуждение

Полученные результаты можно объяснить с точки зрения современных исследований в области обработки естественного языка.

Например, для сущностей «Организация» характерно появление во фразах с большим общим количеством сущностей. Здесь очевидна прямая зависимость, поскольку названия самих организаций очень часто состоят из двух и более слов. Как отмечают Д. Надоу, С. Секине, многие авторы распознают сущность «Организации», идентифицируя повторяющиеся слова в их названиях (Nadeau, Sekine 2007). Например, знание того, что «associates» часто используется в названиях организаций, может привести к распознаванию «Computer Associates» и «Bio-Media Associates» (Gaizauskas et al. 1995), а знание того, что слово «airline» используется в названиях авиакомпаний, может привести к распознаванию «Delta Airlines» (пример из набора данных).

Что касается выделенной зависимости более согласованной оценки экспертами позитивной или нейтральной тональности во фразах с большим общим количеством сущностей и категории «Персона», то это можно объяснить с психолингвистической точки зрения. Обычно процесс анализа настроений и эмоций лучше всего работает с текстом, имеющим субъективный

контекст, чем с текстом, имеющим только объективный контекст. Это связано с тем, что, когда основной текст имеет объективный контекст или перспективу, он отражает некоторые обычные утверждения или факты, не выражая никаких эмоций, чувств или настроения. Субъективный текст содержит информацию, которая обычно выражается человеком, имеющим типичные настроения, эмоции и чувства, что легче поддается классификации экспертами.

Заключение

2В-PLS-анализ показал возможность анализа согласованности оценок тональности фразы экспертами в зависимости от знаний, извлеченных из предложений: именованных сущностей и другой статистики. На согласованность оценок экспертов влияет не только сама категория сущ-

ностей, но и тональность фразы, а также общее количество сущностей и сущностей на латинице во фразе.

В перспективе дальнейшее исследование можно расширить и провести более объемный эксперимент с большим набором данных для более точного выявления латентных зависимостей в текстах с переключением кода для русского и английского языков.

Конфликт интересов

Автор заявляет об отсутствии потенциального или явного конфликта интересов.

Conflict of Interest

The author declares that there is no conflict of interest, either existing or potential.

References

- Agarwal, P., Sharma, A., Grover, J. et al. (2017) I may talk in English but gaali toh Hindi mein hi denge: A study of English-Hindi code-switching and swearing pattern on social networks. In: *9th International Conference on Communication Systems and Networks (COMSNETS)*. Bengaluru: [s. n.], pp. 554–557. <https://doi.org/10.1109/COMSNETS.2017.7945452> (In English)
- Aguilar, G., AlGhamdi, F., Soto, V. et al. (2018) Named entity recognition on code-switched data: Overview of the CALCS 2018 shared task. In: *Proceedings of the Third Workshop on computational approaches to linguistic code-switching*. Melbourne: [s. n.], pp. 138–147. (In English)
- Akischev, T. (2020) The syntax of Kazakh-Russian intrasentential code-switching in reported clauses. *Language. Text. Society*, vol. 7, no. 1. [Online]. Available at: <https://cyberleninka.ru/article/n/the-syntax-of-kazakh-russian-intrasentential-code-switching-in-reported-clauses> (accessed 22.12.2022). (In English)
- Burdygina, M. (2021) Code-switching (Russian—English) in the discourse of IT-specialists from Moscow. *Journal of Siberian Federal University. Humanities & Social Sciences*, vol. 14, no. 10, pp. 1581–1591. <https://doi.org/10.17516/1997-1370-0841> (In English)
- Gaizauskas, R., Wakao, T., Humphreys, K. et al. (1995) University of Sheffield: Description of the LaSIE system as used for MUC-6. In: *MUC6'95: Proceedings of the 6th conference on Message understanding*. [S. l.]: History Publ., pp. 207–220. <https://doi.org/10.3115/1072399.1072418> (In English)
- Kovaleva, V. Yu., Pozdnyakov, A. A., Litvinov, Yu. N., Efimov, V. M. (2019) Otsenka sopryazhennosti morfogeneticheskikh molekulyarno-geneticheskikh modulej izmenchivosti serykh polevok *Microtus* S.L. v gradientnykh usloviyakh sredy [Estimation of the conjugation between morphogenetic and molecular-genetic modules of gray voles *Microtus* S.L. variability along a climatic gradient conditions]. *Ekologicheskaya genetika — Ecological Genetics*, vol. 17, no. 2, pp. 21–34. <https://doi.org/10.17816/ecogen17221-34> (In Russian)
- Krivoshchekov, S. G., Nikolaeva, E. I., Vergunov, E. G., Prihodko, A. Yu. (2022) Multivariate analysis of indicators of inhibitory and autonomic control in orthostasis and emotional situations. *Human Physiology*, vol. 48, no. 1, pp. 20–29. <https://doi.org/10.1134/S0362119721060050> (In English)
- Kuleshova, D. (2020) Code-switching between English and Russian with Russian heritage speakers, born and raised in Russian-speaking families in the USA. *Equity and Access for Language Learners*, vol. 17, no. 1, pp. 44–65. <https://doi.org/10.18060/24431> (In English)
- Luo, T., Chen, S., Xu, G., Zhou, J. (2013) Sentiment analysis. In: *Trust-Based Collective View Prediction*. New York: Springer Publ., pp. 53–68. http://doi.org/10.1007/978-1-4614-7202-5_4 (In English)
- Nadeau, D., Sekine, S. (2007) A survey of named entity recognition and classification. *Linguisticae Investigationes*, vol. 30, no. 1, pp. 3–26. <https://doi.org/10.1075/li.30.1.03nad> (In English)
- Najdenova, K. A., Nevzorova, O. A. (2008) Mashinnoe obuchenie v zadachakh obrabotki estestvennogo yazyka: obzor sovremennogo sostoyaniya issledovaniy [Machine learning for natural language processing: Contemporary State]. *Uchenye zapiski Kazanskogo universiteta. Seriya: Fiziko-matematicheskie nauki — Proceedings of Kazan University. Physics and Mathematics Series*, vol. 150, no. 4, pp. 5–24. (In Russian)
- Naiditch, L. (2000) Code-switching and -mixing in Russian-Hebrew bilinguals. *Studies in Slavic and General Linguistics*, vol. 28, pp. 277–282. (In English)

- Nikolaeva, E. I., Efimova, V. L., Vergunov, E. G. (2022) Integration of vestibular and auditory information in ontogenesis. *Children*, vol. 9, no. 3, article 401. <https://doi.org/10.3390/children9030401> (In English)
- Padmaja, S., Fatima, S., Bandu, S. et al. (2020) Sentiment extraction from bilingual code mixed social media text. In: K. Raju, R. Senkerik, S. Lanka, V. Rajagopal (eds.). *Data Engineering and Communication Technology. Proceedings of 3rd ICDECT-2K19. Advances in Intelligent Systems and Computing*. Vol. 1079. Singapore: Springer Publ., pp. 707–714. https://doi.org/10.1007/978-981-15-1097-7_59 (In English)
- Patra, B. G., Das, D., Das, A. (2018) *Sentiment analysis of code-mixed Indian languages: An overview of SAIL_code-mixed shared task @ICON-2017*. [Online]. Available at: <https://arxiv.org/pdf/1803.06745.pdf> (accessed 22.12.2022). (In English)
- Polunin, D., Shtaijer, I., Efimov, V. (2019) JACOBI4 software for multivariate analysis of biological data. *bioRxiv*. [Online]. Available at: <https://doi.org/10.1101/803684> (accessed 19.11.2022). (In English)
- Rännar, S., Lindgren, F., Geladi, P., Wold, S. (1994) A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. *Journal of Chemometrics*, vol. 8, no. 2, pp. 111–125. <https://doi.org/10.1002/cem.1180080204> (In English)
- Rohlf, F. J., Corti, M. (2000) Use of two-block partial least-squares to study covariation in shape. *Systematic Biology*, vol. 49, no. 4, pp. 740–753. <https://doi.org/10.1080/106351500750049806> (In English)
- Savostyanov, A. N., Vergunov, E. G., Saprygin, A. E., Lebedkin, D. A. (2022) Validation of a face image assessment technology to study the dynamics of human functional states in the EEG resting-state paradigm. *Vavilov Journal of Genetics and Breeding*, vol. 26, no. 8, pp. 765–772. <https://doi.org/10.18699/VJGB-22-92> (In English)
- Singh, V., Vijay, D., Akhtar, S. S., Shrivastava, M. (2018) Named entity recognition for Hindi-English code-mixed social media text. In: *Proceedings of the Seventh Named Entities Workshop*. Melbourne: [s. n.], pp. 27–35. (In English)
- Vergunov, E. G. (2022) Coping space transformation at different levels of university training during the pandemic and the assessment of its integral indicators. *Kompleksnye issledovaniya detstva — Comprehensive Child Studies*, vol. 4, no. 2, pp. 115–123. <https://doi.org/10.33910/2687-0223-2022-4-2-115-123> (In English)