УДК 811.512.122:004.738.5

Комекова С.
учащаяся
Школа-лицей № 120 имени К. Сатпаева
Казахстан, г. Астана
Куракбай М.
ментор образовательной программы ADMIT
Институт науки и технологий Кванджу (GIST)
Южная Корея

РУССКИЕ СЛОВА В КАЗАХСКИХ ОТЗЫВАХ О ПРИЛОЖЕНИЯХ: КАК ИЗМЕРЯТЬ ИСПОЛЬЗОВАНИЕ КАЗАХСКОГО ЯЗЫКА В ЦИФРОВОЙ СРЕДЕ

Аннотация: Казахстан одновременно стремится стать полностью цифровой страной и расширить употребление казахского языка, однако существующие языковые показатели почти не описывают, как казахский используется внутри приложений. Цель статьи – показать на открытых данных, где в цифровой среде казахский язык испытывает наибольшее давление русского, и предложить способ это измерять. Метод: анализ 175 002 открытых казахских онлайн-отзывов набора KazSAnDRA с оценками от одной до пяти звёзд, доля русских слов подсчитана по правилу, построенному по определению первого автора. В отзывах на одну и две звезды русских слов в среднем 3,80%, в отзывах на четыре и пять звёзд – 2,54%, причём русские слова чаще всего стоят поодиночке внутри казахского предложения. Предложено регулярно измерять долю русских слов в отзывах о приложениях как открытый показатель использования казахского языка.

Ключевые слова: казахский язык, языковая политика, переключение кодов, цифровизация, онлайн-отзывы, русские заимствования.

Komekova S.
student
School-Lyceum No. 120 named after K. Satbayev
Kazakhstan, Astana
Kurakbay M.
mentor, ADMIT education programme
Gwangju Institute of Science and Technology (GIST)
South Korea

RUSSIAN WORDS IN KAZAKH APP REVIEWS: HOW TO MEASURE THE USE OF KAZAKH IN THE DIGITAL ENVIRONMENT

Abstract: Kazakhstan aims both to become a fully digital nation and to widen the use of Kazakh, yet existing language indicators say little about how Kazakh is used inside apps. The aim of this article is to show on open data where in the digital environment Kazakh is under the strongest pressure from Russian, and to propose a way to measure it. The method is an analysis of 175,002 open Kazakh online reviews from the KazSAnDRA dataset, rated from one to five stars, with the share of Russian words counted by a rule built from the first author's definition. On average, 3.80% of the words in 1- and 2-star reviews were Russian, against 2.54% in 4- and 5-star reviews, and Russian words mostly stood alone inside a Kazakh sentence. A regular measurement of the share of Russian words in app reviews is proposed as an open indicator of the use of Kazakh.

Keywords: Kazakh language, language policy, code-switching, digitalization, online reviews, Russian borrowings.

Введение

Первый автор замечает, что её друзья часто переходят на русский язык посреди казахской речи, особенно когда не могут вспомнить нужное казахское слово. Кажется, что это происходит чаще, когда человек недоволен, потому что выражать эмоции на русском обычно легче. При этом на русский не переходят полностью: в казахскую речь просто вставляются отдельные слова или фразы.

Казахстан ставит перед собой две цели одновременно. В Послании народу 2025 года Президент Касым-Жомарт Токаев поставил задачу превратить Казахстан в полностью цифровую страну в течение трёх лет [1]. Концепция развития языковой политики на 2023–2029 годы предусматривает рост доли владеющих государственным языком с 81 до 84% и расширение употребления казахского языка в официальных документах и на телевидении [2]. Эти показатели описывают, знают ли люди казахский и использует ли его государство, но почти ничего не говорят о том, что происходит с казахским языком внутри приложений, которыми люди пользуются каждый день, в стране, где русский язык по-прежнему широко распространён [3].

Цель настоящей статьи – на открытых данных показать, где в цифровой среде казахский язык испытывает наибольшее давление русского, и предложить способ регулярно это измерять.

Методы

Использован открытый набор данных KazSAnDRA – отзывы на казахском языке, собранные из магазина приложений, онлайн-библиотеки казахских книг и аудиокниг, картографических сервисов и маркетплейсов [4]. Около трёх четвертей отзывов относятся к мобильным приложениям. Каждый отзыв сопровождается собственной оценкой автора от одной до пяти звёзд, которая показывает, насколько он был доволен. Всего

проанализировано 175 002 отзыва; подробное описание отбора приведено в исследовании первого автора [5].

Русские слова определялись компьютерным правилом, построенным по определению русского слова, которое сформулировала С. Комекова. По каждой группе оценок подсчитывалась доля русских слов, доля отзывов хотя бы с одним русским словом и частота отдельных слов на 10 000 слов. До подсчёта С. Комекова записала прогноз: в отзывах на одну и две звезды русских слов будет втрое больше, чем в отзывах на пять звёзд. Программный код для подсчёта написал ИИ-ассистент, который также проверил правило на выборке размеченных слов.

Результаты

Прогноз подтвердился по направлению, но не по величине. В отзывах на одну и две звезды в среднем 3,80% слов были русскими, в отзывах на четыре и пять звёзд – 2,54%, то есть примерно в полтора раза больше (в 1,49 раза), а не втрое [5]. Разница в основном связана с тем, как часто в отзыве встречается хотя бы одно русское слово: 23,5% отзывов на одну звезду против 15,5% отзывов на пять звёзд.

Большая часть дополнительных русских слов, 62%, – это русские знаменательные слова, среди них существительные, обозначающие то, что не работает в приложении. Слово «реклама» встречалось в отзывах на одну и две звезды примерно в пять раз чаще в расчёте на слово, чем в отзывах на пять звёзд, а слово «обновление» – примерно в семь раз чаще. Эти слова могут быть русскими потому, что технику часто называют по-русски, а не потому, что автор недоволен; имеющиеся данные не позволяют разделить эти причины. Около десятой части дополнительных русских слов составили служебные и усилительные слова вроде «вообще» и «даже»: в группе из 16 таких слов они встречались в отзывах на одну и две звезды в 3,15 раза чаще, чем в отзывах на пять звёзд. Довольные авторы, напротив, благодарили в основном по-казахски: «рахмет» или «рақмет» встречалось

129,8 раза на 10 000 слов в отзывах на пять звёзд, а русское «спасибо» – 0,7 раза.

Русские слова в основном стоят поодиночке. В отзывах на одну и две звезды три из четырёх русских слов не имели рядом другого русского слова, в отзывах на пять звёзд – четыре из пяти. Это согласуется с первым исследованием С. Комековой на тех же отзывах: русский язык обычно входит в казахское предложение одним словом, а грамматика остаётся казахской [6].

Обсуждение

Из результатов следуют три вывода. Во-первых, отдельные русские слова не означают отказа от казахского языка: они встраиваются в казахскую речь, и использовать такие данные нужно с учётом этого. Во-вторых, казахские слова для частых проблем приложений уже существуют, например «жарнама» для рекламы и «жаңарту» для обновления. Данные не показывают, достаточно ли часто люди видят эти слова, чтобы использовать их в момент недовольства, но компании и государственные сервисы с казахским интерфейсом могут употреблять такие слова единообразно и отвечать на жалобы на казахском языке. В-третьих, это можно измерять: отзывы публичны и уже содержат оценку. Подсчёт доли русских слов в отзывах об одних и тех же приложениях во времени, а также до и после изменения казахского интерфейса, показал бы то, чего не показывают существующие показатели: как казахский язык используется в приложениях теми, кто уже на нём пишет.

Ограничения работы. Русские слова определялись компьютерным правилом; проверка показала, что около девяти из десяти отмеченных слов действительно русские, а примерно одно русское слово из шести правило пропускает. Набор собирался как казахские отзывы, поэтому полностью русские отзывы в нём почти отсутствуют, а об авторах отзывов ничего не

известно. Исследование показывает, что низкие оценки и русские слова встречаются вместе, но не доказывает, что одно вызывает другое.

Заключение

Казахстан уже имеет целевые показатели для документов и телевидения. Цифровой стране стоит также отслеживать, как люди пишут по-казахски в приложениях, когда что-то идёт не так, потому что именно там язык используется под давлением. Предложенный показатель – доля русских слов в отзывах с разными оценками – опирается на открытые данные, легко повторяется во времени и может дополнить существующие показатели языковой политики.

Вклад авторов и использование ИИ. Статья основана на исследовании С. Комековой [5]; правило, по которому слова отнесены к русским, построено по её определению и её примерам, вступительный абзац написан ею. Она самостоятельно сверила с опубликованной работой число отзывов и доли русских слов, приведённые в статье, и вручную пересчитала их отношение. М. Куракбай (ментор образовательной программы ADMIT) подготовил текст статьи; программный код исследования и черновик текста подготовлены с помощью ИИ-ассистента (Claude). Оба автора прочитали статью и согласны с ней.

Использованные источники:

1. President Kassym-Jomart Tokayev's State of the Nation Address to the People of Kazakhstan «Kazakhstan in the Era of Artificial Intelligence: Current Challenges and Solutions through Digital Transformation» // Akorda.kz. 2025. URL: https://www.akorda.kz/en/president-kassym-jomart-tokayevs-state-of-the-nation-address-to-the-people-of-kazakhstan-kazakhstan-in-the-era-of-artificial-intelligence-current-challenges-and-solutions-through-digital-transformation-1083029 (дата обращения: 27.10.2026).

2. В Казахстане утвердили концепцию развития языковой политики на предстоящие 6 лет // Zakon.kz. URL: https://www.zakon.kz/pravo/6410885-v-kazakhstane-utverdili-kontseptsiyu-razvitiya-yazykovoy-politiki-na-predstoyashchie-6-let.html (дата обращения: 27.10.2026).

3. Kazakhstan Moves to Promote Its Language Over Russian // The Moscow Times. 06.10.2023. URL: https://www.themoscowtimes.com/2023/10/06/kazakhstan-moves-to-promote-its-language-over-russian-a82689 (дата обращения: 27.10.2026).

4. KazSAnDRA: Kazakh Sentiment Analysis Dataset of Reviews and Attitudes // Proceedings of LREC-COLING 2024. URL: https://aclanthology.org/2024.lrec-main.844.pdf (дата обращения: 27.10.2026).

5. Komekova S. Do unhappy reviewers mix in more Russian? Star ratings and Kazakh-Russian code-switching in 175,002 online reviews // Central Asian Scientific Journal. 2026. № 4(32), т. 1. С. 456–468. URL: https://cajournal.kz/article.php?id=35 (дата обращения: 27.10.2026).

6. Komekova S. Testing classic code-switching constraints on Kazakh–Russian social-media text: an insertional corpus study // Central Asian Scientific Journal. 2026. № 2(30), т. 3. С. 50–58.