Treball de fi de màster de recerca

Verbal borrowings in Russian social media: a diachronic corpus study

Nom i Cognoms Iuliia Lysova

Màster: Lingüística Teòrica i Aplicada

Edició: 2024-2025

Directors: Dra. Gemma Boleda Torrent Dra. Iria de Dios Flores

Any de defensa: 2025

Col·lecció: Treballs de fi de màster

Departament de Traducció i Ciències del Llenguatge

Verbal borrowings in Russian social media: a diachronic corpus study

Iuliia Lysova

Universitat Pompeu Fabra, Barcelona

Abstract

This study examines the adaptation of English-derived verbal concepts related to social media in Russian, focusing on their orthographic, morphological, and semantic integration within online discourse. Based on a diachronic corpus of over 41,000 tweets from 2013 to 2024, the analysis employs computational tools to explore 21 English source concepts, such as 'to post', 'to like', and 'to follow', and their Russian equivalents. The findings reveal that while borrowed forms exhibit greater lemma diversity, their usage frequencies are comparable to those of translated variants. Morphological adaptation shows a growing preference for the suffix *-ну-* in the formation of perfective verbs. The most fully integrated borrowings display a wide range of meaningful affixal variants. Furthermore, the choice and integration of verbal concepts are influenced by a combination of semantic, social, and platform-specific factors, including users' prior experience with other social media interfaces. These findings underscore social media as a dynamic environment for lexical innovation in Russian.

Keywords: borrowings, loanwords, social media, Russian, language change, computational linguistics

Table of contents

1. Introduction

Lexical borrowing, the process by which elements from one language are adopted into another, is one of the most prominent and productive mechanisms of language change. These elements may enter a language to fill lexical gaps, signal prestige, or reflect contact with a dominant linguistic and cultural system. In recent decades, English has become the leading source of lexical borrowing in many languages, including Russian, particularly in domains such as business, technology, and popular culture.

The language of social media offers a unique and evolving context for the integration of English borrowings. Unlike formal registers, digital platforms encourage informal, fast-paced communication and allow users to experiment with new forms, unconstrained by prescriptive norms.

This thesis investigates how English-derived verbal concepts are integrated into Russian within the context of social media discourse, focusing on multiple levels of linguistic adaptation. The study is inspired by a previous project on Persian (Saneei et al., 2023) and is conducted in parallel with two related theses analyzing borrowing in Catalan (Marsó, 2025) and German (Fontanals, 2025). The focus on verbs is intentional: verbs are syntactically more complex, exhibit greater morphological variation, and are often underrepresented in borrowing research, which typically centers on nouns – the most frequently borrowed category.

This work is organized as follows. Section 2 (Background) outlines the central motivations for the study, reviews key literature on lexical borrowing, particularly in Russian, and presents the research objectives and questions. Section 3 (Methodology) describes the data collection, processing procedures, and analytical methods employed. Section 4 (Results) presents the findings and addresses the research questions in light of the data. Finally, Section 5 summarizes the main findings and offers concluding remarks.

2. Background

One widely cited definition of borrowing is offered by Hoffer (2002), who describes it as "the process of importing linguistic items from one linguistic system into another, a process that occurs any time two cultures are in contact over a period of time". Lexical borrowing usually refers to the adoption of individual words or groups of words, also known as 'loanwords'.

The Russian language has a long history of lexical borrowing, shaped by cultural, political, and economic interactions. Historically, Russian absorbed words from various languages, including Scandinavian, Turkic and Germanic languages, as well as Greek, and French, particularly during periods of significant cultural exchange (Janurik, 2010; Yorova, 2017). However, since the mid-20th century, English has emerged as the dominant source of borrowings, driven by the global influence of the United States and the rise of globalization. The dissolution of the Soviet Union in 1991 opened Russia to Western cultural and technological trends, leading to an influx of anglicisms in media, technology, and youth culture. This also made American English the primary source of new lexical items in Russian (Krysin, 1996; Dyakov, 2003; Janurik, 2010). In recent decades, English borrowings have become especially prominent in Russian internet and social media language, where users enjoy high linguistic freedom and prioritize creativity over traditional norms.

As for the motivations behind lexical borrowing, scholars generally agree that borrowings arise from the need to express new concepts, enhance prestige, or resolve linguistic gaps – such as avoiding taboo expressions or homonymy (Krysin, 1996; Dyakov, 2003; Janurik, 2010). Haspelmath (2009) adds that these motivations are not always strictly linguistic; in many cases, languages possess the internal resources to create new words, yet speakers still opt for loanwords. He distinguishes two main types: **cultural borrowings**, which introduce terms for novel or foreign concepts (e.g., Russian *компьютер* [komp'uter] for 'computer'), and **core borrowings**, which replace or coexist with existing native terms – often due to the prestige of the donor language (e.g., Russian *сервис* [servis] for 'service', coexisting with the native *обслуживание* [obsluzhivanie], which may carry slightly different stylistic or contextual connotations (Krysin, 1996; Haspelmath, 2009).

2.1. Types of lexical borrowings in Russian

Borrowings can be classified according to various criteria, the most fundamental of which concerns the type of adaptation a source word undergoes when integrated into the recipient language. As Haspelmath (2009) notes, the linguistic properties of the donor language often do not align with those of the recipient language, requiring certain modifications for successful integration. This dimension of integration primarily involves form and meaning – that is, phonetic, orthographic, morphologic and semantic adaptation. Its goal is to reshape the word to fit the phonological and structural constraints of the recipient language while retaining its original meaning.

In the case of Russian, lexical borrowings can be broadly divided into two main types: **(1) direct borrowings, where the foreign stem is incorporated through transliteration or transcription,** and **(2) translations, where native roots are used to reproduce the meaning of the original term – either as calques or semantic extensions**. Each of the types will be defined below.

Most borrowings from English undergo phonetic and orthographic adaptation in Russian due to the differences between the two script systems – Latin and Cyrillic. **Transliteration** transfers the graphic form in a character-by-character manner (e.g., 'listing' → *листинг* [listing], 'slogan' → *слоган* [slogan]), while **transcription** is a way to capture phonetic form (e.g., ‘know-how’ → *ноу-хау* [nou-hau], ‘office’ → *офис* [ofis]) (Timofeeva, 1995).

However, current orthographic guidelines for borrowings in Russian offer limited support to users, as they are largely based on established norms for already assimilated words (Nechaeva, 2011). Orthographic dictionaries, including the most notable Orthographic Dictionary by Lopatin, published by the Vinogradov Institute of the Russian Language (Lopatin, 2007), occasionally provide inconsistent or contradictory spellings for newly borrowed words. As a result – but most importantly due to users’ autonomy in adopting new lexical items, usually without consulting any guidelines – orthographic adaptation tends to exhibit considerable variation. Two major factors contributing to this variability are: (1) the lack of direct equivalents for certain English vowel sounds in Russian, and (2) the inconsistent treatment of doubled consonants in borrowed stems.

When the Russian speakers attempt to integrate new concepts with the help of native lexical and semantic resources they will usually resort to the translation with either calques or semantic extensions. **Calques** involve the literal translation of English terms into Russian, preserving their structure or meaning. In the XX century, structural calques were more prominent and included examples like *небоскрёб* ([neboskryob] from ‘skyscraper’, with *небо* for ‘sky’ and *скрёб* for ‘scraper’) (Dyakov, 2017). **Semantic calquing**, the key subtype in the XXI century, entails a Russian word acquiring a new meaning by analogy with a foreign word’s sense. For example, the Russian word *кейс* [keis] for ‘brief-case’ has adopted additional business connotations (‘business case’) mirroring English usage (Gorbov, 2015).

Semantic extension occurs when an existing Russian word acquires a new meaning under the influence of a foreign language without being a direct translation of the corresponding term. For example, during the COVID-19 era, the lexeme *корона* [korona] for ‘crown’ emerged as a neological homonym for ‘coronavirus', used informally with a lowercase letter (Radbil, 2021).

Figure 1 presents a modified version of the diagram used in Marsó (2025). It was adapted for Russian, summarizing the classification of loanwords employed in this study.

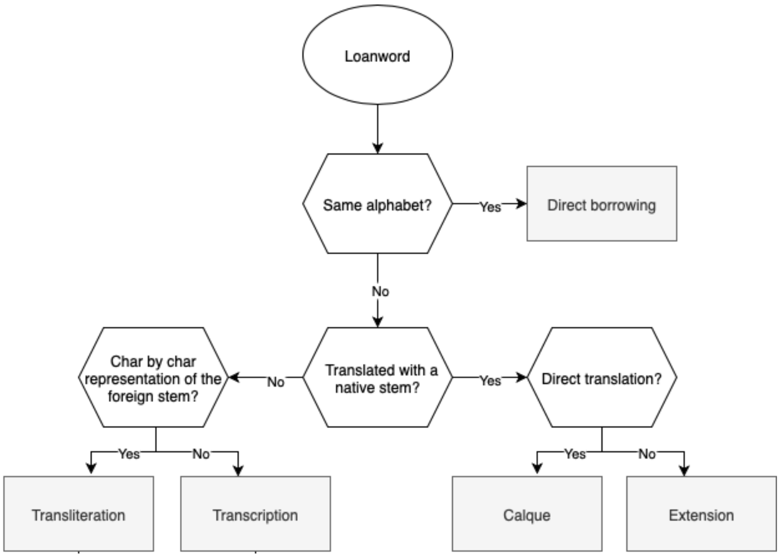

Figure 1. Classification of borrowings from the point of view of the form (first layer)

According to previous research, direct borrowings dominate over translations, probably due to their immediacy (Dyakov, 2017; Gorbov, 2015). In the context of social media, additional factors may influence the type of adaptation users adopt. For example, users may interact with either English-language interfaces or Russian-localized versions, where localizers typically favor native lexical items. Consequently, the interface language to which users are exposed may affect their lexical choices alongside other sociolinguistic factors.

2.2. Morphological adaptation of verbal borrowings in Russian

Loanwords often require adaptation not only in form but also in grammar to function within the recipient language (Haspelmath, 2009). In languages like Russian, where verbs have conjugational and aspectual distinctions, borrowed items must be assigned to appropriate inflectional and syntactic categories in order to be grammatically usable. This highlights a second dimension of the borrowing process: morphosyntactic integration.

Recent research confirms that most borrowings from English in Russian, and slang words in particular, undergo conjugation and declination processes (Vorobyeva, 2009; Lackova, 2021). Verbs borrowed from English are typically adapted through affixation, acquiring Russian inflectional markers and aspectual pairs and being assigned to one of the productive conjugation classes.

2.2.1. Aspectual adaptation

Loan verbs, especially anglicisms, typically do not display aspectual distinctions in their source language. However, in Russian verbs are inherently marked as either perfective or imperfective. This usually can be done with prefixes, such as *с-, за-*, *на-*, *про-, по-* etc. (*делать* [delat'] (IMPF) – ***с****делать* [sdelat'] (PERF) for 'to do'), or suffixes, such as *-ну-* (*прыгать* [prigat'] (IMPF) – *прыг****ну****ть* [prignut'] (PERF) for 'to jump'). The challenge arises from the fact that the formation of aspectual pairs in Russian is not fully morphologically regular, particularly for recently borrowed or neologistic verbs (Olsson, 2021).

Some earlier linguistic theories argued that only suffixed perfective verbs (unlike their prefixed counterparts) form genuine aspectual pairs with imperfectives. This view was based on the observation that prefixes in Russian often convey not only aspectual but also lexical meaning, and therefore cannot be considered "pure" markers of aspect. For example, the prefix *на-* in ***на****писать* [napisat'] 'to write' is used to form a perfective to *писать* [pisat'] 'to write', but in ***на****печь* [napech'] 'to bake a lot of something' the prefix also adds the meaning 'a lot' to the imperfective base form *печь* [pech'] 'to bake'. However, more recent research (e.g., Janda & Lyashevskaya, 2011) showed that both types of aspectual pairs exhibit similar grammatical behavior in actual usage, and should therefore be regarded as equally valid. In regards to what prefix to choose, the so-called "overlap hypothesis" (Janda et al. 2013) claims that the choice of aspectual affix depends on the meaning of the verb and that the meanings of the verb and the prefix usually overlap.

At the same time, there is a possibility that one aspectual affix is the "regular form" in modern Russian for new borrowings adoption and therefore used for almost all new aspectual pairs. Previous corpus-based studies have identified the prefix *за-* as the most productive means of perfectivization in colloquial Russian (Gjervold, 2013). In contrast, an experimental study by Olsson (2021), which focused on the creation of new verbs from light verb constructions with borrowed nouns, found that Russian speakers more often favored the suffix *-ну-* for forming perfective verbs. This inconsistency in the results keeps the question of the preferred way for perfectivization of newly adopted verbs open.

2.2.1. Derivational capacity

Acquiring semantic autonomy is a key stage in the integration of borrowed lexical items into the Russian language system. As Bogoslavskaya (2015) notes, it is at this point that a borrowing begins to "settle" into the lexicon of the recipient language, acquiring clear semantic boundaries and developing its own paradigmatic relationships. Also, high derivational capacity may be a sign of high level of adaptation and semantic autonomy.

In Russian, this process is often signaled by the ability of a borrowed verb to participate in native word-formation processes, particularly aspectual derivation, as was mentioned before, and prefixation in general. For example, the verb *постить* [postit'] ('to post') has produced forms like *запостить* [zapostit'] ('to post [once]') and *напостить* [napostit'] ('to post excessively'), each reflecting a specific nuance of digital action. Through such processes, borrowed verbs become embedded within broader semantic networks, aligning with native Russian patterns of prefixation and aspectual differentiation, and thus achieving full lexical and semantic integration.

At the same time, high morphological variation is not always indicative of a high degree of integration. In some cases, such variation may signal instability, where different morphological choices do not convey distinct meanings but rather reflect a lack of user consensus. For example, in the case of ‘to repost’, early forms could include the transliterated ***ре****постить* [repostit’], the transcribed ***ри****постить* [ripostit’], and the partially calqued ***пере****постить* [perepostit’], illustrating inconsistency rather than stable integration into the language.

2.3. Constructions with supporting (light) verbs

According to Wohlgemuth (2009), many languages integrate borrowed verbs using two main strategies: **direct insertion** and the **light verb strategy**. In the direct insertion approach, the borrowed verb stem is incorporated directly into the morphological system of the recipient language, undergoing inflection like native verbs. In contrast, borrowed nouns may also appear in constructions, where they remain mostly uninflected and combine with a semantically weak native verb (light verb) that carries the grammatical information of the predicate (‘to put a like’).

Russian clearly follows the first strategy, regularly adapting borrowings through native inflectional mechanisms. At the same time, Russian also exhibits a version of the light verb strategy. In these cases, however, the borrowed element typically appears as an inflected noun (e.g., *поставить лайк* [postavit’ laik] ‘to put a like’, where ‘like’ is in the Accusative case).

In the linguistic literature on Russian, there is no unified approach to defining light verbs. They are often referred to as “supporting verbs” and may include expressions such as *нести ответственность* [nesti otvetstvennost'] 'to bear responsibility', where the verb carries minimal semantic weight (Maiko, 2022). The primary criterion for identifying such constructions is the loss or reduction of the verb's original lexical meaning. For example, in the expression *написать пост* [napisat' post] ('to write a post'), the verb *написать* does not typically refer to the physical act of writing, but rather conveys the action of posting. In such constructions, the noun serves as the semantic core, while the verb functions primarily as a grammatical carrier. These verbs may not be "light" in the strict sense but exhibit a reduced semantic load relative to their original lexical usage.

In Russian, these constructions are favored in formal discourse for their precision and ability to incorporate modifiers, as in *проводить тщательный анализ* [provodit' tschatel'niy analiz] ('to conduct a thorough analysis') compared to the less flexible *тщательно анализировать* [tschatel'no analizirovat'] ('to thoroughly analyze') (Blinova, 2024).

This phenomenon has also been studied in the domain of social media in other languages. In their study of English-Persian bilingual constructions, Saneei et al. (2023), demonstrating the increasing use and evolution of such structures in Persian social media, often prevailing over simple verbs derived from borrowed stems. However, in Russian there is no research of this phenomena in the context of social media. It is therefore important to investigate whether constructions with supporting verbs are used by Russian-speaking social media users, influenced by a highly globalized and internationalized environment, and how they compare in frequency and function to fully lexicalized borrowed verbs.

2.4. Research questions

This study investigates how English-derived verbal concepts related to social media are adapted and developed in Russian, particularly within the informal and dynamic context of online discourse. While borrowing and loanword integration have been widely studied in Russian linguistics, existing research often presents inconsistent findings and rarely involves diachronic corpus analysis based on large-scale naturalistic data. This study focuses specifically on verbs that represent platform-specific actions, such as 'to post', 'to follow', 'to tweet', or 'to like', which are frequently borrowed from English and exhibit varying degrees of orthographic, morphologic, and semantic integration**.**

To address this overarching topic, the following questions are explored:

1. What strategies do social media users prefer when selecting adaptation types and making orthographic and morphologic adaptation decisions? Do these align with the previous research and to what extent?
2. Which borrowed verbal concepts show the highest degree of linguistic adaptation, and what sociolinguistic or technological factors contribute to their integration?
3. How do borrowed verbs compete with or coexist alongside constructions with supporting verbs, and what does this reveal about the dynamics of verbal integration in contemporary Russian?

The study anticipates that adaptation will be uneven across dimensions. Users are expected to favor using borrowed stems and transcription for their orthographic adaptation. Verbs with borrowed stems are likely to be fully integrated into the Russian morphological system, particularly through aspectual and derivational patterns, with a probable preference for the suffix *-ну-* in perfectivization. The verbs with higher derivational capacity are expected to align with the most popular platform actions, while constructions with supporting verbs are likely to persist alongside lexicalized forms, especially for newer or less conventionalized borrowings, though their number may decrease with the adaptation process of the borrowed concept.

3. Methodology

To answer the questions outlined in the Background section, the study employs a range of computational linguistic tools. These include automated web scraping to build a corpus of Russian-language tweets, large language models (LLMs) for semantic annotation, and Python-based analysis for frequency counts, verb form tracking, morphological analysis and visualization. Morphological and tagging tools are also used to capture the productivity and variation of borrowed verb forms and to create custom inflectional dictionaries[1].

3.1. Terminology

The following terminology is used throughout the thesis:

Lemma – the base form of a word that is used to represent all the other possible forms, infinitive for verbs. Verbs with prefixes derived from those without prefixes are considered to be separate lemmas, also it is the case for verbs of different aspects. For example, *постить* [postit'], *запостить* [zapostit'] and *постнуть* [postnut'] are three different lemmas for 'to post'.

Word form – a specific grammatical or inflected form of a lemma, including variations by tense, person, number. For instance, *постит* [postit] is a present-tense, third-person singular verb form of the lemma *постить* [postit'] ('to post'). Sometimes referred to as 'verb form', when applied to verbs.

Type – a unique lemma or word form occurring in the corpus. Each distinct word form or lemma is counted once as a type, regardless of how many times it appears. For example, if *запостил* [zapostil] ('posted') appears 50 times, it still counts as one type of the word form.

Token – an individual instance of a lemma or a word form in the corpus. All occurrences of the same lemma or word form are counted separately as tokens. Using the previous example,

[1]code and data can be found at: https://github.com/julialysova/language_change_rus

запостил appearing 50 times would be counted as 50 tokens.

3.2. Word choice and wordform dictionaries

For the purposes of analysis, a set of 21 English-language social media-related concepts was selected. These concepts are commonly used in online discourse, denoting elements of social media interface or actions that can be produced within this medium. In English, they can appear either as verbs or as nouns. In Russian, they have at least one borrowed variant (transliterated or transcribed). The selected items for both verbal and nominal concepts included 'ban', 'block', 'comment', 'follow', 'friend', 'like', 'message', 'mute', 'pin', 'post', 'reply', 'repost', 'restrict', 'retweet', 'share', 'tag', 'tweet', 'thread', 'unfollow'. For light verb constructions, nouns such as, 'story', 'follower' were considered in addition to the mentioned ones.

For each concept, a corresponding expression in Russian or several were identified, drawing on the author's linguistic competence and familiarity with current usage in social media contexts (e.g. – *лайк* [laik], *лайкать* [laikat'] for 'like'; *тегнуть* [tegnut'], *тэг* [teg], *тагать* [tagat'], *отметить* [otmetit'] for 'tag'; *блок* [blok], *заблокировать* [zablokirovat'], *блокнуть* [bloknut'] for 'block', etc.). This list was refined and expanded iteratively throughout the analysis as new forms emerged in the data. In total, 164 Russian lemmas denoting the studied concepts (both verbs and nouns) were collected before data scraping (explained in the Section 3.3) and 19 additional lemmas were registered during the analysis.

To enable the extraction of all possible grammatical realizations of borrowed verbs from the dataset, a comprehensive list of inflected word forms was generated using a **Python** library **pymorphy2**[2] morphological analyzer (Korobov, 2015). This tool was chosen for its extensive coverage of Russian morphology, including support for both standard and

[2] Can be found at: https://github.com/kmike/pymorphy2

non-standard inflectional paradigms – crucial for capturing the full range of variation in borrowed verb usage. The list of inflected forms was designed to include all theoretically possible forms for each verb under study. The automatically produced forms were then manually reviewed to correct conjugational errors and to ensure the inclusion of less frequent or non-standard variants, particularly those likely to appear in informal social media discourse. All the word forms then were manually annotated with their English equivalent, lemma, aspect, and adaptation type and subtype according to its borrowing strategy: transliteration, transcription, calque, or semantic extension according to the classification provided in the Background section. Lemmas that were found later in the analysis were also added to the dictionaries with a 'no' in the column "Was scrapped" to be analyzed separately from the scraped forms. Table 1 and 2 show the structure of the dictionaries for verbs and nouns correspondingly.

English word	Russian lemma	Russian word form	POS	Aspect	Adapt. type	Adapt. subtype	Adaptation degree	Was scrapped
ban	банить	баню	VERB	IMPF	borrowed	transliterated	full	yes
block	блокнуть	блокнул	VERB	PERF	borrowed	transcribed	full	yes
comment	комментировать	комментировала	VERB	IMPF	translated	calque	full	yes
follow	подписаться	подпишусь	VERB	PERF	translated	extension	full	yes
follow	фолловиться	фолловятся	VERB	IMPF	borrowed	transliterated	full	no

Table 1. Selection of rows from the inflectional dictionary. There is at least one example of each feature

English word	Russian lemma	Russian word form	POS	Adapt. type	Adapt. subtype	Adaptation degree	Preposition	Was scrapped
ban	бан	бан	NOUN	borrowed	transliterated	full	yes	yes

Table 2. Illustration of the dictionary for nouns. It contains an additional column "Preposition", which means whether the noun needs a preposition to build a light verb construction (as in добавить в друзья [dobavit' v druzia] 'to add a friend' meaning 'to friend')

3.3. Corpus data preparation

We selected X.com (formerly Twitter) as the data source, since it was also used in Saneei et al. (2023) along with Instagram. Additionally, X.com is widely used and characterized by short, user-generated texts, which offers access to informal language use and lexical patterns in real-time communication.

The data was collected using the Twitter (X.com) Search Scraper, a tool available through the Apify.com platform[3] that allows scrapping of large amounts of data with some customizable parameters for better precision. The scraping process targeted Russian tweets and all lexical items from the predefined list in their Russian lemma forms, along with two to three of their most frequent inflected forms, as the search operates on exact string matches rather than lemmatized forms or substrings. Overall, the search comprised 24 words in Latin script (English concepts to account for any possible code switches), 164 Russian lemmas denoting these concepts and additionally 320 word forms for them. Later 3 concepts (‘subscribe’, ‘direct message’, ‘profile’) were excluded because no borrowed variants were scrapped for them neither in verb nor in noun forms.

We targeted four days – January 24, April 24, July 24, and October 24 – of the years 2013, 2018, 2023, and 2024 spanning an 11-year period. To minimize repetition and more formal discourse, retweets with added content and tweets from news accounts were excluded. The extraction process was carried out by a research assistant at the university due to time limitations. A total of 139,720 tweets were scrapped fo Russian. From all the data that the scrapper extracts for each tweet, we keep the columns that are relevant: the id of the tweet (string of numbers), the content of the tweet (only the text), username (a string of text) and the timestamp of the tweet (the date of publication). However, not all of the tweets contained the target stems, as comments of the associated tweets could also be included. Thus, further filtering was needed.

Tweets underwent a multi-stage filtering and processing pipeline. First, all the tweets were deduplicated by the id and the text in case the tool scrapped identical tweets. Then, they

[3] The tool can be found at: https://apify.com/web.harvester/easy-twitter-search-scraper

were deduplicated by the user and the text, which aimed to prevent cases, when a user posted a number of tweets with identical content (may be considered as spam). This deduplication gave us 106,907 tweets. Then the tweets were lowercased and filtered by the presence of target stems (for example, *фоллов* [follov] for *фолловить* [follovit'] 'to follow') derived from the selected lexical items to account for all the word forms, giving us 85,429 tweets. Second, tweets were pre-processed to remove non-linguistic content – including hashtags, user mentions, URLs, and other extraneous symbols – using **regular expressions**, a standard and efficient method for text pattern matching and cleaning. Additionally, the scrapping tool sometimes included tweets in different languages, because they also used Cyrillic script, though the number of them was not high. To account for them anyway, all the tweets that contained special characters that appear in other Cyrillic languages, such as Mongolian, and are not used in Russian were removed, which decreased the amount of the tweets to 84,862. Third, additional language identification was conducted using the **SpaCy** natural language processing library (Honnibal & Montani, 2017) and their *ru_core_news_lg* model; only tweets classified as Russian were retained and at this step the dataset amounted to 80,236 tweets. All the dates appeared in the UTC format, so it was decided to additionally extract the year and put it to a separate column.

In the final stage, wordform extraction was carried out. We decided not to use the columns with the search query, since the scrapped word could appear in the deleted hashtags and links. Also, sometimes tweets contained more than one target word and extracting all of them was a more efficient way of using the data.

The pre-processed tweets were tokenized using **SpaCy**. However, lemmatization was intentionally not performed using SpaCy, as its lemmatizer is not always reliable for rare, non-standard, or newly borrowed verbs – precisely the kinds of forms under investigation in this study. Instead, each token was compared against the comprehensive list of inflected forms generated with **pymorphy2**, ensuring that all grammatical realizations of the target verbs were captured directly. Matching word forms were recorded in a structured format: each identified form was placed in a separate column adjacent to the original tweet, allowing a single tweet to appear multiple times in the resulting dataset if it contained multiple relevant word forms. Tweets that did not contain target words were eliminated. Overall, 72,065 tweets were obtained, comprising 92,559 instances of different word forms of target lemmas.

3.4. Detection of social media senses

Given that a major part of target verbs in the dataset are polysemous and may occur with both social media-specific and general meanings (e.g., *отметить* [otmetit'] can mean both 'to tag' and 'to celebrate'), each instance of a target verb form identified in the corpus required semantic disambiguation and annotation of whether the word was used in a social media-related sense or in a non-social media context.

Due to the size of the dataset, full manual annotation was not feasible. Therefore, a semi-automated two-step annotation strategy was implemented:

1. Annotation of a training dataset of 4,000 rows with Large Language Models (LLMs), informed by recent findings showing the effectiveness of LLMs and their ensembles in particular for classification tasks (Zhen, 2024).
2. Use of the training dataset for fine-tuning of a **RuBERT classification model**[4] (Kuratov & Arkhipov, 2019). RuBERT is a model based on BERT architecture that was trained on the Russian part of Wikipedia and news data. It was chosen for its strong performance on Russian-language NLP tasks and broad adoption in linguistic research.

This approach not only allowed to avoid manual annotation of the whole dataset while keeping high accuracy, but also reduced cost for using a paid API's account for LLMs access.

[4] The model can be found at https://huggingface.co/DeepPavlov/rubert-base-cased

3.4.1. Annotation with an ensemble of LLMs

The LLM-based annotation framework followed a three-step pipeline present in Figure 2.

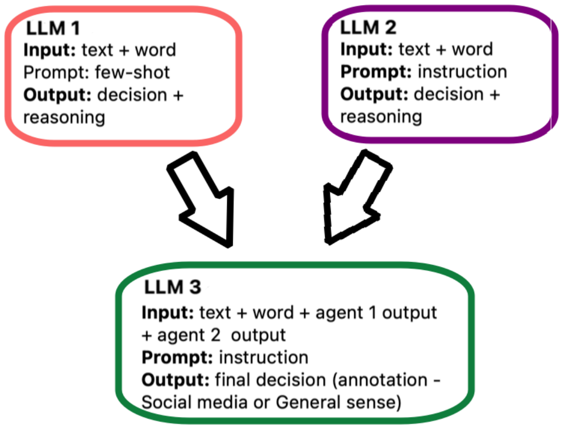

Figure 2. Architecture of the LLMs ensemble for social media senses annotation showing what kind of input and output are expected for each of the participating LLM

To interact with the models, we used a technique known as prompting, in which a text string (the prompt) is provided to a language model to guide its output (Jurafsky & Martin, 2023). The model generates a continuation conditioned on the prompt, effectively using it as context to produce task-relevant responses. In this study, we employed two types of prompting: few-shot and zero-shot. In few-shot prompting, the prompt includes a small number of labeled examples to illustrate the task and improve performance. In contrast, zero-shot prompting provides only task instructions without examples, relying entirely on the model's understanding of the context (see Table A1 in the Appendix for prompts and models). Different prompts techniques were chosen to add variability to the models' answers and reduce probable bias associated with identical prompts.

LLM 1: Received the tweet and target word form as input, along with a few-shot prompt, and was instructed to determine whether the verb was used in a social media sense, providing a brief justification.

LLM 2: Received the same input with a zero-shot prompt performing the same classification task and offering its own justification.

LLM 3: Functioned as an adjudicator, taking into account both prior annotations and their justifications to make the final decision.

For each of the agents a different model was used (Table A1). The models were accessed via **Groq-hosted API**, which offered fast access to the variants, without the need to use own computational resources. This multi-LLM pipeline was orchestrated using **LangChain**[5], an LLM application framework that allows modular and transparent coordination of multiple models.

This ensemble method achieved 94% agreement with a manually annotated gold-standard set of 200 randomly selected tweets and 97% accuracy on a 100 tweet validation set. In contrast, each individual model on its own achieved between 86%–93% accuracy, confirming the 4-8% improvement gained through the ensemble approach.

Model name	Number of parameters	Accuracy 200	Accuracy 100
Llama 4 Scout	17 billion	0.835	0.87
Llama 3.3	70 billion	0.865	0.93
Llama 4 Maverick	17 billion	0.86	0.9
Ensemble	**3 models 17 + 17 + 70**	**0.94**	**0.97**

Table 3. Accuracy the tested models using the training set of 200 and 100 tweets. Ensemble approach showed the best accuracy

The training dataset of 4,000 tweets was annotated by the approach with two labels: “Социальные сети” (Social media) and “Общий смысл” (General sense) that were then translated to binary labels “1” and “0” accordingly. The training dataset contained 2,394 (~60%) rows that were assigned with “1”, meaning slightly unbalanced classes. The optimal number of tweets for training was selected iteratively – starting with 1,000 and incrementing by 1,000 until reaching nearly 90% at the second stage of the annotation process.

[5] More information can be found here: https://www.langchain.com/

3.4.2. RuBERT fine-tuning

The training dataset was used for further fine-tuning of the RuBERT model. Fine-tuning is a transfer learning technique that adapts a pretrained language model, like BERT, to a specific downstream task (in our case – classification of the social media context) by training it further on a smaller, labeled dataset with minimal adjustments to the original model's parameters (Jurafsky & Martin, 2023).

Input texts were formatted as sentence pairs, where Sentence A contained the original tweet and Sentence B contained the classification question (e.g., 'Is the word *X* used in a social media sense?'). This formulation aligns with the Question answering approach, when the model is given a context and a question for it. All the hyper parameters were selected iteratively, based on what gives the best performance of the model (Table A2 in the Appendix).

The fine-tuned model achieved accuracy of 89.6 and **F1 score of 89.5**, a metric that balances precision and recall, which is crucial given potential class imbalance in the dataset (Jurafsky & Martin, 2023). The fine-tuned RuBERT model was then applied to annotate the remainder of the corpus (88,559 rows). For the model training results see Table A3 in the Appendix.

After annotation of tweet-word pairs, the final dataset comprises 41,693 unique tweets with 55,689 word forms classified as used in a social media-related sense. For now, this includes both verbs and nouns.

3.5. Analysis

The analytical phase of this study focused on identifying patterns of adaptation of English-derived social media verbal concepts in Russian in terms of adaptation types, as well as on phonological, orthographical and morphological levels. The analysis was both quantitative and qualitative in nature and relied on corpus-based methods supported by computational tools (mainly Python-based scripts in Google collab and VS Code**).**

First, all verbal forms annotated as social media-related were analyzed for their adaptation type. The token frequency and lemma diversity of each adaptation subtype (transcription, transliteration, calque, and extension) were calculated across the four sampled years of data. Frequencies were computed both per type and individually for each target English verb.

To account for variation in concept popularity that could skew distribution (i.e., more popular actions appearing more frequently), we used normalized counts. For each year, and for each English verb, we calculated the proportion of Russian verb forms belonging to each adaptation type by dividing the number of tokens of that type by the total number of Russian tokens corresponding to the same English verb. These per-word proportions were then summed across all English verbs to obtain the total normalized frequency for each adaptation type per year.

$$NormalizedCount_{i,y} = \sum_{w} R_{i,y,w} = \sum_{w} \frac{T_{i,w,y}}{N_{w,y}}$$

$R_{i,y,w}$ = relative frequency of adaptation type *i* for word *w* in year *y*, $T_{i,w,y}$ = number of tokens of adaptation type *i* for English verb *w* in year *y*, $N_{w,y}$ = total number of Russian tokens for English verb *w* in year *y*

This way, each English verb contributes equally to the overall distribution, regardless of its raw frequency.

Second, the orthographic adaptation of borrowed verbal stems was examined, focusing particularly on the transcription of vowels and the representation of doubled consonants. A subset of verbal stems corresponding to phonologically ambiguous English sounds – such as /æ/ and /e/ – was analyzed. Stems variants such as *шер*, *шэр*, and *шар* for ‘share’, and *тег*, *тэг*, and *таг* for ‘tag’, were automatically extracted from the target word forms found in the dataset and counted (overall and per year). Patterns of orthographic choice were compared against the results of previous research (Nechaeva, 2011). Additionally, the stems frequency distributions were examined diachronically to see whether there is a shift of usage of certain stems over time.

Third, morphological adaptation of borrowed forms was assessed through an inventory of derived verb forms, focusing on the range of productive prefixes and suffixes attached to borrowed stems (e.g., ***за***_лайкать_ [zalaikat’], ***от***_лайкать_ [otlaikat’], _лайк_***нуть*** [laiknut’] for ‘to like’). For each target English verb, unique Russian lemmas and prefix-suffix combinations of their word forms were extracted to measure morphological and semantic productivity.

This step also included aspect creation analysis, as aspects are also formed with prefixes and suffixes. We focused on the distribution of perfective and imperfective forms across *the most frequent* Russian perfective and imperfective lemma types used for each target English word. Particular attention was given to the distribution of affixes as *за*-, and -*ну*- to check its alignment with existing research on aspect formation (e.g., Janda & Lyashevskaya, 2011; Olsson, 2021). Additionally, diachronic analysis of the usage of *за*-, and -*ну*- for perfective verb forms creation was done to look at the dynamics of the choice among users of X.com.

Finally, the role of constructions with supporting verbs in the borrowing process was explored by identifying all instances where English-derived nouns (e.g., *лайк* [laik] (‘like’), *пост* [post] (‘post’), *сторис* [storis] (‘stories’)) occurred as direct or indirect objects in multiword expressions with semantically reduced Russian verbs (e.g., *поставить лайк* [postavit’ laik] (‘to put a like’), *выложить сторис* [vilozhit’ stroris] (‘to post a story’)).

To extract the constructions, all texts were tagged for part of speech using spaCy. For each tweet, occurrences of the target nouns were identified, along with any verbs appearing in their immediate context: at position -1 or +1 (as Russian verbs may appear either before or after the direct object), and at position +2 to account for cases where the object is separated from the verb by a pronoun or preposition (only for nouns that allow preposition in this position). Cases, when the gap between the light verb and the noun is more than 1 word were not covered due to the time constraints, because a more complex syntactic analysis would be needed. Each verb found in these positions was compared against a list of supporting verbs that was compiled by the author and iteratively refined throughout the analysis. The list contained 46 verb lemmas, including light verbs or verbs with reduced semantic load, like *сделать* [sdelat'] ('to do'), *написать* [napisat'] ('to write') and *выложить* [vilozhit'] ('to upload'). If a verb matched an item in this list, it was recorded in a separate column in the dataset for further analysis.

All the found constructions were manually reviewed for correctness. The main criterion was to include phrases where the noun serves as the semantic core, while the verb carries reduced lexical meaning. Ideally the phrase should be easily substituted with a lexicalized verb (like *запостить* [zapostit'] 'to post') without loss of meaning. Then, selected phrases were counted automatically to find the most frequent nouns and light verbs to compare their distribution with the lexicalized verbs for each English concept.

4. Results

A total of 41,693 tweets, spanning four years within an 11-year period, were analyzed. The Figure 3 shows a twofold decrease in tweet volume over time. While this decline is likely influenced by the official ban of X.com in Russia in 2022 (which now requires users to access the platform via VPN) the downward trend was already evident by 2018. This suggests that the declining activity may also reflect a broader reduction in engagement with the platform among Russian-speaking users, independent of legal restrictions.

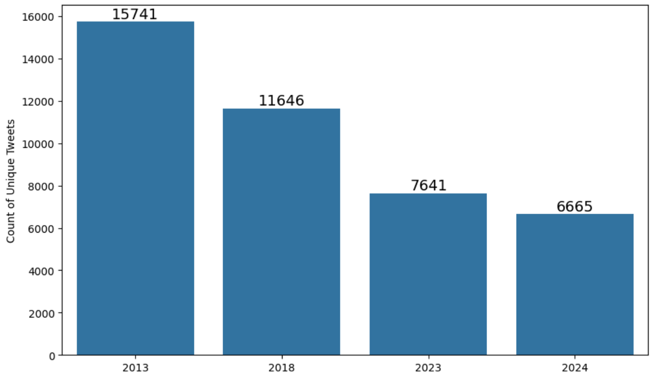

Figure 3. The distribution of unique tweets counts with social media-related content across years after filtering, preprocessing and annotation

In the corpus, we have found 22,095 instances of different verb forms representing 96 Russian verb lemmas that correspond to 18 target English verbs. One verb – 'message' – did not have any lexicalized Russian verb equivalents in the corpus.

4.1. Adaptation types

In our data, 11 of the 18 studied English verbs are adapted in at least two different ways. Usually they appear both as translated verbs and the verbs with a foreign stem (hereinafter referred to as "borrowings" or ¨borrowed lexemes¨), but 6 are found only as the borrowed ones.

Figure 4 demonstrates the distribution of unique lemma types across adaptation strategies. Transcribed lemmas, such as *лайкать* [laikat'] ('to like'), account for 49% of all unique lemmas, consistent with findings from previous studies (Dyakov, 2012; Nechaeva, 2011).

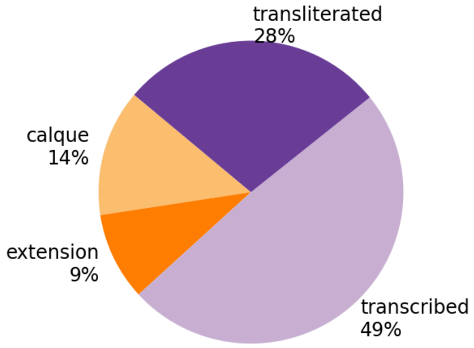

Figure 4. Percentage of total counts of all the Russian lemma types found in the corpus per their adaptation subtype

However, the overall frequency of transcribed word forms is comparable to that of the calqued ones, such as *блокировать* [blokirovat'] ('to block'). This can be seen from Figure 5 showing the normalized diachronic counts. This finding suggests that many transcribed forms are used infrequently and some variants may still be undergoing stabilization. Figure 5 also reveals an approximately even distribution of borrowings and translations across years. Only a slight decrease in calqued forms and an increase in extensions usage can be observed for the translated forms, while borrowed verbs show relative consistency across years.

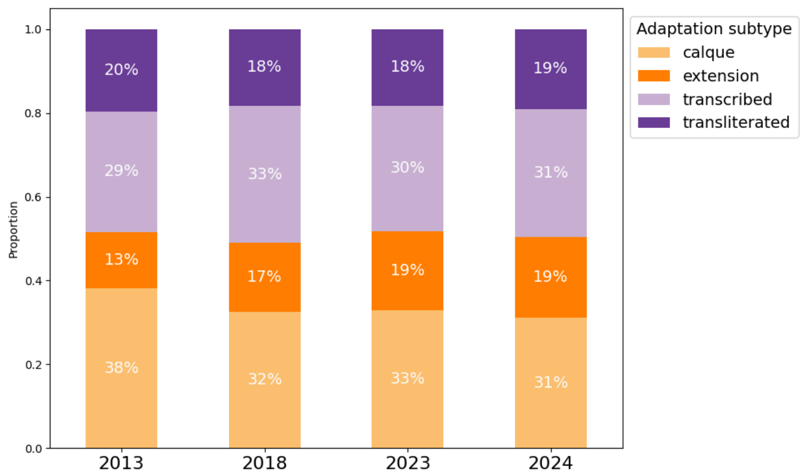

Figure 5. Normalized counts of Russian verb forms occurrences across time, showing the distribution of the preferred adaptation subtypes by the users of [X.com](X.com) for the target social media concepts

Figure 6 below shows the distribution of adaptation subtypes for each English concept individually.

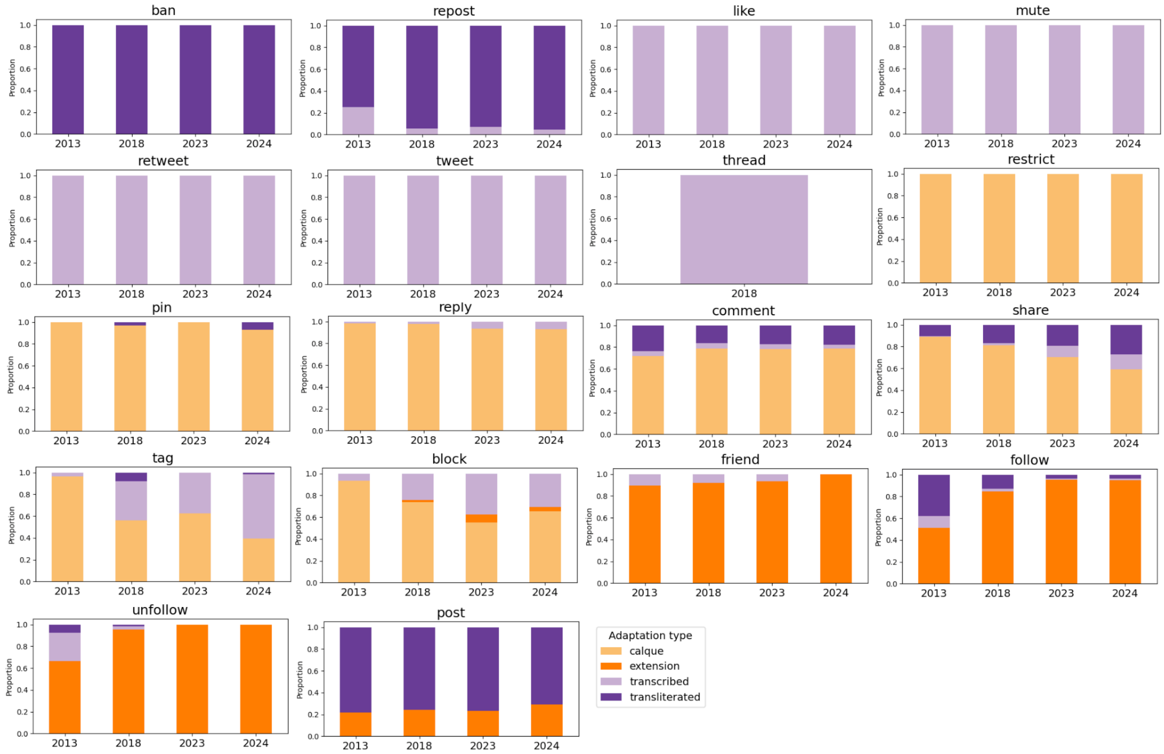

Figure 6. Percentage of the Russian verb form counts with different adaptation subtypes. The distribution is shown per English word with the darkest color corresponding to the most frequent subtype

Remarkably, most verbs show consistency across time and are used predominantly in a single adaptation type. Seven verbs that appear in the corpus exclusively as borrowings, such as *твитить* [tvitit’] (‘to tweet’), do not have any instances of other adaptation types. For ‘to tweet’, ‘to retweet’, ‘to like’, and ‘to mute’ direct Russian equivalents are either lacking, semantically insufficient, or inconvenient. Supposedly, users tend to utilize verbs with borrowed stems more when no existing Russian verb fully captures the meaning of the original concept.

At the same time, some concepts – such as ‘to tag’, ‘to share’, and ‘to block’ – show increasing use of borrowed forms despite having widely used native equivalents from the outset. This trend challenges the notion that borrowings emerge primarily for reasons of lexical immediacy. Instead, it suggests that stylistic preference and semantic precision also contribute to the appeal of borrowed forms, especially when native equivalents (e.g., *отметить* [otmetit'] for 'to tag', or *поделиться* [podelit'sya] for 'to share') have more ambiguous meanings.

A case of a favorable translation is the verb 'to follow', which is predominantly expressed through the extended form *подписаться* [podpisat'sya] ('to subscribe'), a usage likely influenced by the localized interface of platforms such as YouTube. This translation conveys a related, though not identical, meaning and is preferred over the borrowing *фолловить* [follovit'] or a more direct calque such as *следовать* [sledovat'], which is actually never used in this context. This suggests that sometimes users may adopt semantically adjacent native terms that feel familiar and functionally appropriate, rather than introducing new equivalents.

Overall, the integration of English-derived verbs into Russian is shaped by a complex interplay of semantic adequacy, user familiarity, social conventions, and platform-specific terminology. While borrowings usually fill lexical gaps, they also may replace less precise native terms. Meanwhile, some translated forms, initially introduced through localized interfaces on other platforms, may be extended to analogous functions on new platforms, reflecting continuity in user behavior

4.2. Orthographic adaptation

The high number of transcribed forms seen in 4.1 may also be a result of variation in phonetic and orthographic adaptation from the Latin script to the Cyrillic alphabet. And, as mentioned in the Background section, high variation may be a sign of lower levels of adaptation.

In particular, the English vowels /æ/ (as in 'tag' and 'ban') and /e/ (as in 'share') do not have direct one-to-one equivalents in the Russian vowel system. As a result, users employ different strategies when transcribing these sounds, typically using the letters *е*, *э*, or *а*. Nechaeva (2011), based on observations of previously assimilated borrowings, notes that transcribed variants using *е* or *э* are generally preferred, and that over time, stems initially spelled with *э* tend to shift to *е* as adaptation progresses, unless otherwise specified by established orthographic dictionaries.

Figure 6 shows the distribution of the stems across the analysed period. For ‘tag’, we find that stems with *е* are more commonly used, which aligns with Nechaeva's observation and also with the fact that the Orthographic Dictionary by Lopatin lists only the form *тег* [teg]. As shown in Figure 7, the process of orthographic assimilation has reached its advanced stage over the analyzed period, with a shift from *э* to *е* in the spelling of the stem.

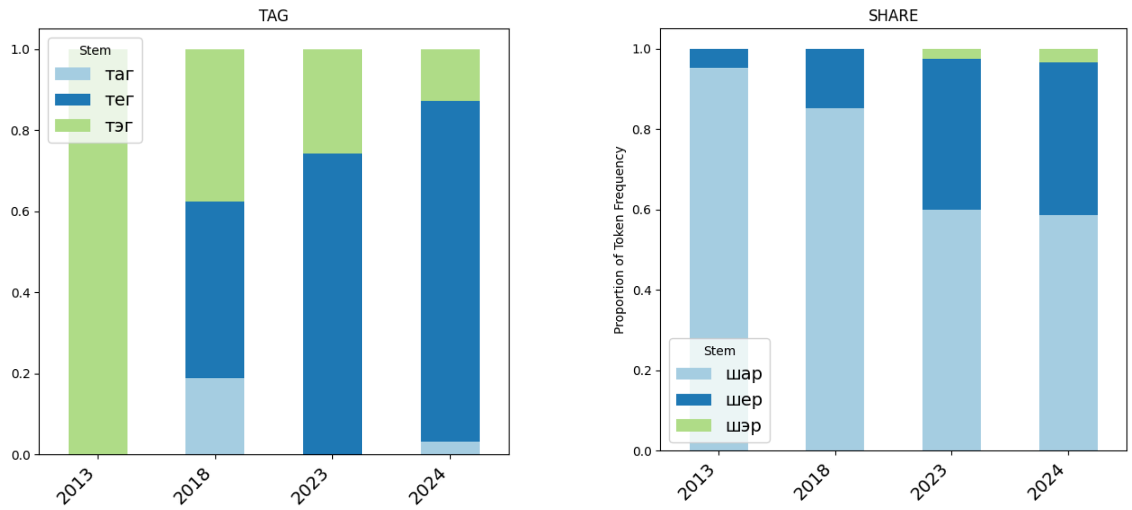

Figure 7. Percentage of stems with е, э, or а in Russian borrowings of the verbs ‘share’ and ‘tag’ (based on all the verb form tokens of borrowings)

As for ‘to share’, users exhibit a strong preference for stems with *а* (Fig. 7), though there is some evidence of change: the proportion of stems with *е* is gradually increasing over time. This may indicate an ongoing process of assimilation for ‘share’. The abundance of transliterated stems with *а* could be attributed to the fact that such words were often borrowed from written social media contexts, without reference to their original pronunciation. Transliteration may also serve as a strategy to avoid the uncertainty associated with transcription (Nechaeva, 2011). However, transliterated variants may never evolve to their transcribed versions. For example, for ‘to ban’, the Orthographic Dictionary by Lopatin lists only the transliterated variant *банить* [banit'], which also appears exclusively in this form in the analyzed corpus, suggesting a high level of convention and consequently assimilation of the transliterated form.

Consonant variation also arises when the English stem contains duplicated letters, as native Russian words typically feature double consonants only at morpheme boundaries, but not within a stem, as in *ll* in 'follow' or *mm* in 'comment'. While this is not common to native lexemes, users often preserve both consonants in borrowed stems, particularly in the early stages of assimilation (Nechaeva, 2011). Over time, some forms may lose the duplicated letter, though the factors influencing this shift remain unclear. Figure 8 shows that in the analyzed corpus, Russian verb stems for 'follow' and 'comment' occur more frequently in their doubled forms; however, for 'follow', there is a slight increase in the simplified form *фолов* [folov], though its overall proportion remains small. This trend is not observed for 'comment', which may be due to the fact that the borrowed verb *комментить* [kommentit'] is perceived as a shorter and more informal variant of the already fully assimilated *комментировать* [kommentirovat'], which appears in dictionaries and retains the double *м* in the stem.

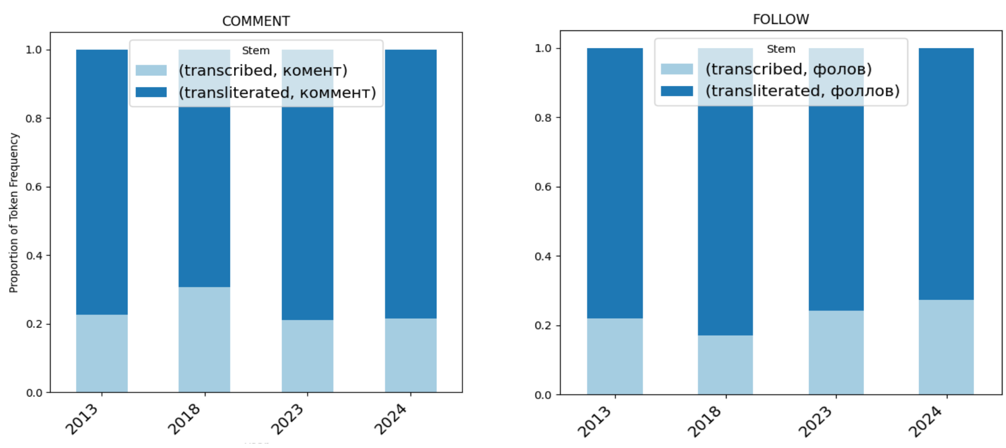

Figure 8. Percentage of stems with single or double consonant in Russian borrowings of the verbs 'comment' and 'follow' (based on all the verb form tokens of borrowings)

Overall, users tend to favor transliteration for doubled consonants. However, vowel adaptation, specifically /æ/ and /e/, shows less consistency and weaker adherence to any certain strategy, reflecting greater variation in user preferences and instability.

4.3. Morphological adaptation

All the borrowed verbs examined in this study are morphologically integrated into Russian, adopting native affixes and fitting into the inflectional system.

4.3.1. Adaptation to the aspectual system

In our data, forms in the perfective aspect are more frequent, and the number of unique lemmas in this aspect is nearly twice as high as in the imperfective (81 vs. 41 respectively), suggesting a greater morphological productivity in the perfective aspect. This tendency may reflect either the action-oriented, result-focused nature of social media interactions or the high variation in the ways of creating the perfective aspect. However, it remains unclear whether this distribution is specific to social media discourse or reflects a broader pattern in the Russian language. Further investigation is needed to determine whether corpora from other domains exhibit similar distributions of aspectual verb forms.

As shown in Figure 9, we find a high variation in how often the verbs appear in one or the other aspect. As of 2023-2024, among productive English verbal stems, 14 out of 18 produce the Russian borrowings in both perfective and imperfective aspectual forms. Two verbs – ‘to thread’ and ‘to restrict’ – do not appear in these years. The remaining two verbs (‘to unfollow’ and ‘to pin’), which appear only in a single aspect, belong to the less frequent group within the corpus. This may signify their lower levels of adaptation or less interest in these actions by users of X.com.

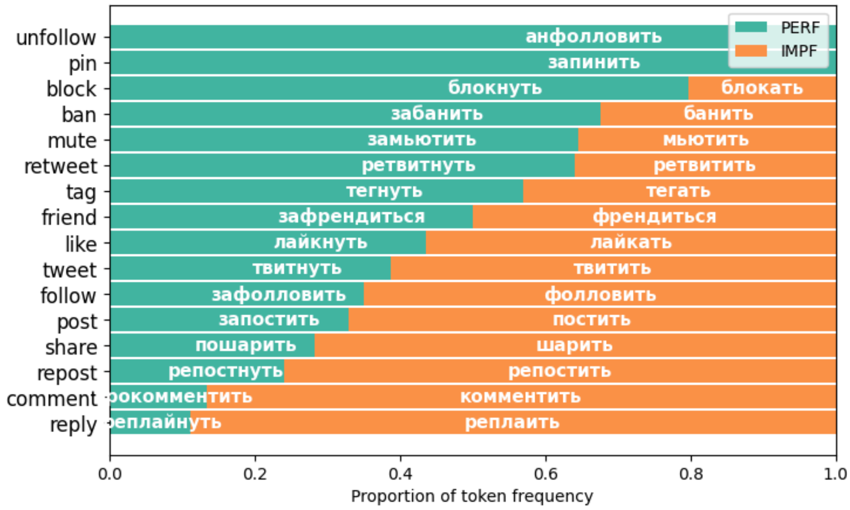

*Figure 9. Percentage of Russian verb forms tokens with **different aspects** per English verb and the most frequent lemmas per English verb and per aspect indicated in the white font (2023-2024)*

Prevalence of a certain aspect for each concepts seems to be connected with innate episodicity of the word – the actions that are usually isolated, like ‘to ban someone’ appear more in the perfective, while actions that can be repetitive – ‘to tweet’, appear as imperfectives more frequently.

As for the most preferred strategy for forming the perfective aspect, among the most frequent perfective lemmas over the whole corpus it is the use of the prefix *за-*. However, the suffix *-ну-* also appears as a common and productive means of perfectivization (see also Table A6 in the Appendix). At the same time, considered across time, two clear trends can be observed – the prefix *за-* loses its productivity over time, giving the place to the suffix *-ну-,* as shown in Figure 10.

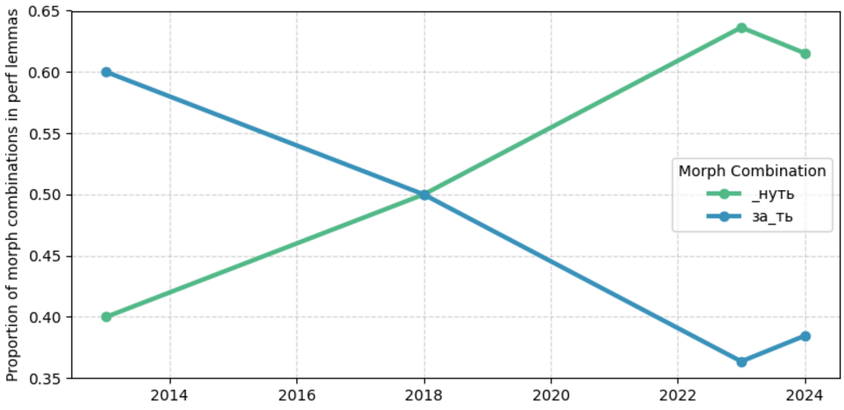

Figure 10. Trend of за- and -ну- for perfective verbs creation. Counted as a proportion between the two affixes among the most frequent perfective Russian lemmas for each English word per year.

These results retrospectively coincide with the observations of Gjervold (2013) that found the prefix *за-* was the most productive strategy in 2010s for forming the perfective aspect in Russian, however the current trends show consistency with Olsson's (2021) findings that *-ну-* maybe a more preferable strategy for contemporary Russian nowadays. The exact reasons for this shift are not entirely clear, but it is likely that a new generation of X.com users is developing its own preferences for perfective formation, with the suffix *-ну-* perceived as more natural or expressive in contemporary digital discourse.

4.3.2. Derivational capacity

In terms of derivational capacity – that is, the ability to generate new lexical units – a useful measure is the diversity of prefix-affix combinations that a borrowed stem can support.

For this analysis, all word forms containing the target foreign stems were examined in the corpus to identify additional forms not initially known to the author. This search revealed 56 additional word forms across 19 previously unregistered lemmas. Then they were analyzed together with all the scraped forms. As Figure 11 shows, the results indicate that verbs such as 'to like', 'to retweet', 'to post', and 'to ban' exhibit the greatest morphological variation in their adapted Russian forms, suggesting a higher potential for expressing nuanced meanings in user-generated discourse.

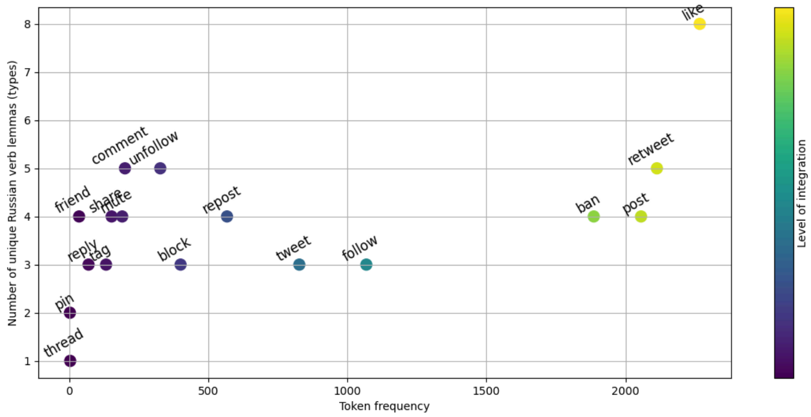

Figure 11. Level of derivational integration illustrated as the interaction of the token frequency of the Russian verb forms and the number of their unique affix combinations. Lighter colors mean higher levels of adaptation

The set of Russian lemmas derived from ‘to like’ illustrates a high degree of morphological and semantic productivity (Table A7 in the Appendix). These forms incorporate a variety of Russian verbal prefixes (*от-, по-, про-, за-, на-, пере-, до-*) and aspectual suffixes (*-нуть, -ать, -ануть*). Lemmas like *полайкать* [polaikat’] (‘to like a little bit’)*, отлайкать* [otlaikat’] (‘to put likes on everything’)*, пролайкать* [prolaikat’] (‘to put like successively’)*, залайкать* [zalaikat’] (‘to put too many likes’)*,* allowing speakers to express nuanced variations in action.

High derivational capacity is also shown by the Russian lemmas derived from ‘to ban’ that form at least 4 different shades of senses (see also Table A7). Interestingly, in recent years, the word ‘to ban’, represented by its Russian equivalents, gained an exceptional popularity among X.com users, winning over ‘to follow’ and ‘to like’ (Figure A4 in the Appendix). This may be attributed to the influx of bots to the platform and high polarization of opinions caused by recent social events, which makes people ban users with different or propagated points of views.

At the same time, in Figure 12 we can observe that words of relatively lower frequency, such as ‘unfollow’, can also have a high number of different affix combinations.

The borrowed lemmas for this English word include *отфолловливать* [otfollovlivat’], *отфолловиться* [otfollovlivatsa], *отфолловить* [otfollovit’], *анфолловить* [anfollovit’], *расфолловить* [rasfollovit’], where the last three have the same meaning (‘to unfollow’ in perfective aspect), they just bear different perfective prefixes. Over time, the number of lemmas decreases to just one, along with a general decline in the frequency of borrowed forms for ‘unfollow’. This trend may reflect a lack of consensus about the word among users and a lower degree of integration of this borrowing into the Russian language, as discussed in the Background section.

Figure 12. Lifespan of lemmas derived from borrowed stems, illustrating varying degrees of morphological adaptation. Stable forms across the period (‘like’) may indicate a high level of integration, while cases with multiple derivatives early on and only one remaining toward the end of the period suggest morphological instability and reduced integration

4.3.3. Verbs made from nouns

Interestingly, concepts that exist only as nouns in English may develop verbal forms in Russian. For example, in the corpus we could find the noun ‘story’ to appear not only as a borrowed noun but also as a derived verb – *сторить* [storit’] (IMPF) (‘to post stories’). This illustrates the morphological flexibility of Russian, where users actively verbalize borrowed nouns to fit syntactic needs. A more systematic analysis of such cases is left for future research.

4.4. Constructions with supporting verbs

A total of 991 constructions with supporting verbs were analyzed. It is a remarkably lower amount compared to the number of lexicalized verbs. The majority of these constructions involve nouns with borrowed stems, and notably, only one of them incorporate nouns written in their original Latin script – 'like' (Table A7 in the Appendix). It appeared several times in 2013 and only once in 2018. This suggests that code-mixing in user-generated content may occur at the earlier stages of the word adoption.

Figure 13 shows that the nouns that most frequently appear in supporting verb constructions are related to the English words 'post' (*пост* [post]), 'comment' (*комментарий* [kommentarii]), 'like' (*лайк* [laik]), and 'tweet' (*твит* [tvit]).

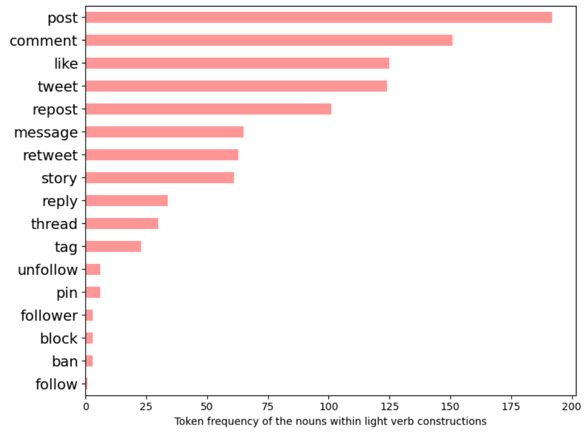

Figure 13. The most frequent concepts (per token frequency of all their Russian equivalents) appearing in the verbal constructions extracted from the corpus

Figure 14 below shows proportions between lexicalized verbs and supporting verb constructions usage. It is obvious that most concepts tend to be adapted as fully inflected verbs (e.g., *мьютить* [miutit'] 'to mute'). However, some concepts appear exclusively in supporting verb constructions – for instance, *написать сообщение* [napisat' soobschenie] 'to write a message'. When examining individual trends, most verbs either maintain a stable level of supporting verb construction usage or show a slight increase over time, as with 'retweet' and 'reply'. This challenges our assumption that light verb constructions are used to introduce new borrowed concepts to the language and decrease over time. Instead, the data suggest that certain concepts develop stable usage patterns over time, either analytical or morphological – whichever the best align with their morphological and semantic characteristics.

Figure 14. Percentage of Russian verb forms usage (blue) and percentage of Russian light verb phrases usage (red) per English word

In terms of productivity, there appears to be a correlation between the frequency of a borrowed noun and the variety of verbs it combines with. Figure 15 demonstrates that *пост* [post] ('post'), *твит* [tvit] ('tweet'), комментарий [kommentariy] ('comment'), *сторис* [storis] (‘stories’) are of the most frequent ones and emerge as the most productive items, occurring with the widest range of light verb constructions. More details can be found in Table A8 in the Appendix.

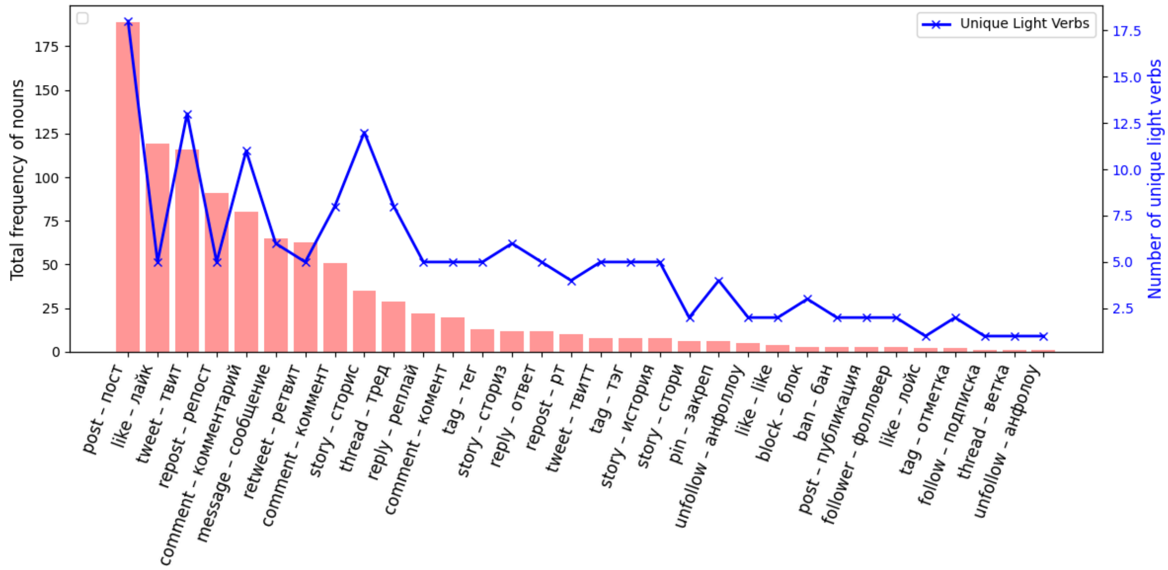

Figure 15. Frequencies of Russian nouns within light verb constructions and the number of unique light verb lemmas they use (per Russian noun lemma). The plot shows dependency of the noun frequency to the variation of light verbs used with it

However it is worth mentioning that supporting verb constructions seem to be a less productive option for social media users, for whom morphological verb creation is a more preferable strategy.

5. Discussion and conclusion

This thesis has presented the analysis of adaptation of English-derived social media-related verbal borrowings in Russian. While there is a long-standing tradition of studying borrowing in Russian, much of the existing literature either focuses on mass media data or explores slang on social media in general. This study contributes to closing that gap by examining borrowing patterns in a real-world, informal setting and specifically analyzing verbs, which have received relatively less attention in previous research than nouns. We believe that the conducted research has helped us to answer the research questions outlined in the Background section.

5.1. Preferred strategies of adaptation

One of the most notable findings is the predominance of borrowings over translations in terms of lemma diversity along with nearly equal usage frequencies of both types over time. Interestingly, some verbs that were initially more common in their calqued forms show a shift toward borrowed forms as the prevailing type. This supports the idea that direct incorporation of foreign stems (through transcription and transliteration) is not only a convenient strategy for rapid initial lexical integration, but also it can be driven by stylistic preferences and semantic precision.

Regarding the users' choice of orthographic adaptation strategy, the data reveals a complex relationship between traditional norms and new borrowings. While users generally choose transliteration for double consonants, vowels show a wider variation. Users utilize *тег* [teg] for ‘tag’, but favor *а* in *шар* for 'share', despite the sound /e/ in ‘share’ is more closed than /æ/ in ‘tag’. However, the temporal trends show that the words, especially ‘share’, may still be in the process of adaptation (Nechaeva, 2011).

The morphological adaptation patterns, instead, demonstrate remarkable consistency with previous research. The dominance of the prefix *за-* for perfectivization in the 2010s and the productive use of *-ну-* suffixation in 2020s demonstrate that preferred ways of morphological adaptation evolve over time and users are still in the search for the most natural or expressive forms. Importantly, aspectual integration takes place across nearly all borrowed verbal concepts, despite the absence of aspectual distinctions in the source language. Moreover, the choice of aspect often aligns with the inherent semantics of the borrowed action.

5.2. Factors contributing to linguistic adaptation success

The second research question concerning which verbal concepts achieve highest linguistic adaptation reveals that frequency and conceptual necessity drive integration success. Verbs like *лайкать* [laikat'] ('to like'), *банить* [banit'] ('to ban'), and *твитить* [tvitit'] ('to tweet') demonstrate the richest morphological productivity, generating diverse prefix-suffix combinations that express nuanced semantic distinctions. This productivity correlates strongly with the absence of direct Russian equivalents, supporting Haspelmath's (2009) cultural borrowing hypothesis where novel concepts necessitate lexical innovation. This also suggests that semantic adaptation of a borrowing is more successful when it fills specific functional gaps, rather than merely offering stylistic alternatives to existing vocabulary. For instance, the borrowed verb *фолловить* [follovit'] ('to follow') has seen limited adoption compared to the native-derived *подписаться* [podpisatsa], a calque of 'to subscribe'. Over time, *подписаться* has come to represent the broader concept of following across various social media platforms, demonstrating how existing usage patterns and prior interface localizations can shape the integration and success of new borrowings.

Also, the popularity of certain concepts may be influenced by social settings, as it is the case with the word 'ban' that became more popular over time probably due to major social events.

5.3. Verb constructions vs. lexicalized verbs

The third research question confirms that constructions with supporting verbs in contemporary Russian are less popular than in some other languages. Contrary to Saneei et al.'s (2023) findings about light verb constructions in Persian social media, Russian users demonstrate a clear preference for lexicalized verbs over the constructions (22,095 vs. 991 instances).

This preference suggests that Russian's rich verbal morphology remains highly productive in digital contexts. This is further confirmed by verbalization of nouns ('story' → сторить [storit’] (‘to story’)), which demonstrates Russian speakers' active usage of morphological resources.

5.4. Limitations and future work

While the current study addressed the research questions posed at the outset, its scope is not without limitations. The dataset, although sizable (over 41,000 tweets), remains constrained by the number of time points (only 4 days for each year) and the selection of target items. Some lexical items were underrepresented because they were not included in the scrapping (specifically verb lemmas derived from nouns), and their adaptation patterns remain difficult to evaluate. Future work should expand the corpus both temporally and lexically, allowing for a more robust diachronic analysis and the emergence of broader trends. Also, other social media platforms can be examined to see whether these results are platform-specific. Despite these limitations, the findings suggest that social media is a great way to look at how borrowed verbs are integrated into Russian, both grammatically and semantically. Further exploration of how social media-related concepts are integrated in the general discourse can deepen our understanding of contemporary Russian as it continues to evolve in many other domains.

References

Blinova, O. V. (2024). Attributive and adverbial modifiers in Russian light verb constructions (based on the Law Corpus data). *Saint Petersburg State University. (In Russian).*

Bogoslovskaya, V. R., Dolgenko, A. N., & Kosyreva, M. S. (2015). From “selfie” to «селфи»: Pathways of adaptation of anglicisms in contemporary Russian. *Izvestiya VGPU*, (9–10), 125–129. *(In Russian).*

Dyakov, A. I. (2003). The reasons for the intensive borrowing of anglicisms in modern Russian. *Language and Culture*, (2003), 35–43. *(In Russian).*

Dyakov, A. I. (2012). Anglicisms: Borrowing or word-formation. *Philological Sciences: Theoretical and Practical Issues*, (5), 72–76. http://www.gramota.net/materials/2/2012/5/17.html *(In Russian).*

Dyakov, A. I. (2017). Calqued anglicisms in the Russian language. *Obrazovanie i nauka v sovremennykh realiiakh*, *2*(1), 26–35. https://doi.org/10.21661/r-462201 *(In Russian).*

Vorobyeva, S. V. (2009). Grammatical assimilation of the recent anglicisms in Russian. *Herald of the Minsk State Linguistic University. Series 1: Philology*, (5), 178–186. *(In Russian).*

Fontanals, G. (2025). Study on social media loanwords in German [Unpublished master’s thesis]. *Universitat Pompeu Fabra.*

Gorbov, A. A. (2015). On semantic calquing and 'secondary borrowing' in Russian at the turn of the 21st century. *Voprosy Jazykoznanija*, (1), 87–101. *(In Russian)*

Gjervold, J. (2013). Prefix variation: A comparison between slang and contemporary standard Russian. *Poljarnyj Vestnik, 16*, 33–46.

Haspelmath, M. (2009). Lexical borrowing: concepts and issues. Loanwords in the world's language: A Comparative Handbook. 35-54. 10.1515/9783110218442.

Hoffer, B. L. (2002). Language borrowing and language diffusion: An overview. *Intercultural Communication Studies*, *11*(4), 1–37.

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. https://spacy.io

Janda, L. A., & Lyashevskaya, O. (2011). Aspectual pairs in the Russian National Corpus. *Scando-Slavica*, *57*(2), 201–215. https://doi.org/10.1080/00806765.2011.626611

Janda, L. A., Endresen, A., Kuznetsova, J., Lyashevskaya, O., Makarova, A., Nesset, T., & Sokolova, S. (2013). Why Russian aspectual prefixes aren't empty: Prefixes as verb classifiers. *Bloomington: University of Tromsø.*

Janurik, Sz. (2010). The integration of English loanwords in Russian: An overview of recent borrowings. *Studia Slavica*, *55*(1), 45–65. https://doi.org/10.1556/SSlav.55.2010.1.3

Jurafsky, D., & Martin, J. H. (2023). Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models (3rd ed.). https://web.stanford.edu/~jurafsky/slp3/

Korobov, M. (2015). Morphological Analyzer and Generator for Russian and Ukrainian Languages. In: *Khachay, M., Konstantinova, N., Panchenko, A., Ignatov, D., Labunets, V. (eds) Analysis of Images, Social Networks and Texts. AIST 2015. Communications in Computer and Information Science, vol 542. Springer, Cham.* https://doi.org/10.1007/978-3-319-26123-2_31

Kuratov, Y., & Arkhipov, M. (2019). Adaptation of deep bidirectional multilingual transformers for Russian language. *arXiv preprint*, arXiv:1905.07213. https://arxiv.org/abs/1905.07213

Krysin, L. P. (1968). *Inoyazychnye slova v sovremennom russkom yazyke* [Foreign-language borrowings in modern Russian]. Moscow: Nauka. *(In Russian).*

Lacková, M. (2021). Morphological peculiarities of lexical units of English origin in contemporary Russian slang: Dictionary and corpus analyses. *RUDN Journal of Language Studies, Semiotics and Semantics, 12*(3), 632–651. https://doi.org/10.22363/2313-2299-2021-12-3-632-651

Lopatin, V. V. (Ed.). (2007). A Russian Orthographic Dictionary. Moscow: Russian Academy of Sciences. T*he V.V. Vinogradov Russian Language Institute.* Accessed through https://orthographical.slovaronline.com/

Maiko, T. (2022). Support verb constructions: A Russian-Italian contrastive analysis. *Firenze University Press*. https://doi.org/10.36253/978-88-9273-983-3

Marsó, L. (2025). Piular or tuitejar? A study of social media loanwords in Catalan [Unpublished master's thesis]. Universitat Pompeu Fabra.

Nechaeva, I. V. (2011). Current issues in the orthography of foreign borrowings. *Moscow: Azbukovnik Publishing Center. (In Russian).*

Olsson G. (2021) How Recently Borrowed Verbs in Russian Form Perfective Aspect: An Experimental Approach. *Slověne,* Vol. 10, № 1, p. 392–413. https://doi.org/10.31168/2305-6754.2021.10.1.17

Saneei, S., Shojaei, R., & Borghesani, V. (2023, October 24-26). Syntactic-Semantic Analysis of the Emergence of Novel English-Persian Bilingual Complex Predicates [Poster presentation]. Society for the Neurobiology of Language, Marseille, France. https://www.neurolang.org/program/presentation/?id=1753

Radbil, T. (2021). New Russian lexis in the aspect of cultural appropriation of loan words. *European Proceedings of Social and Behavioural Sciences, 109*, 421–432. https://doi.org/10.15405/epsbs.2021.09.46

Timofeeva, G. G. (1995). New English borrowings in the Russian language: Spelling and pronunciation. *Saint Petersburg: Yuna.*

Yorova S. N. (2017) The history of borrowing words in the Russian language. In: *Scientific notes of Khujand State University. Humanities*. No. 2 (51).

Zhen Tan, Dawei Li, Song Wang, Alimohammad Beigi, Bohan Jiang, Amrita Bhattacharjee, Mansooreh Karami, Jundong Li, Lu Cheng, and Huan Liu. (2024). Large Language Models for Data Annotation and Synthesis: A Survey. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 930–957, Miami, Florida, USA. Association for Computational Linguistics.* https://aclanthology.org/2024.emnlp-main.54.pdf

Appendix

Table A1. Models and prompts used for the LLM annotation

LLM	Model	Prompt
1. Few-shot	*meta-llama/llama-4-maverick-17b-128e-instruct* Multimodal, instruction-tuned model with 17 billion parameters and 128 experts. An instruction-tuned model is a large language model that has been specifically trained to follow instructions. This is achieved through a process called instruction tuning, where the model is fine-tuned on a dataset of instructions and their corresponding outputs.	**System message:** Ты - интеллектуальный лингвистический помощник. **Human message:** Твоя задача - определить, в каком контексте используется указанное слово - в контексте социальных сетей или в общем смысле. Примеры: Текст: жизнь женщины би лайк (be like) Слово: лайк Ответ: Общий смысл Т: Когда можно лайкать и репостить моих выдр С: лайкать О: Социальные сети Т: Сейчас я расскажу вам историю, от которой кровь стынет в жилах. С: историю О: Общий смысл Т: ты писала мне грязные пристающие комментарии под нюдсы. извините, я тогда не могу удержаться не зафолловить тебя. С: комментарии О: Социальные сети Т: я выложила новую историю в инстаграм. С: историю О: Социальные сети Т: Александр Васильевич занял пост Главного директора С: пост О: Общий смысл Прежде чем ответить, коротко поясни свой выбор. Теперь посмотри на этот текст:{text} Слово: {word} Ответ:
2. Zero-shot	*meta-llama/llama-4-scout-17b-16e-instruct* Multimodal, instruction-tuned model with 17 billion parameters and 16 experts.	**System message:** Ты - интеллектуальный лингвистический помощник. **Human message:** Твоя задача - определить, в каком контексте

		используется указанное слово - в контексте социальных сетей или в общем смысле. Посмотри на твит: "{tweet}" И анализируемое слово: {word} Теперь определи, используются ли это слово применительно к соцсетям или в своем общем значении. При решении опирайся на слова из контекста твита, не додумывай ничего сам. Отвечай "Социальные сети" если слова используются применительно к соцсетям, и "Общий смысл" если нет и приведи свои доводы.
3. Judge	*llama-3.3-70b-versatile* State-of-the-art multilingual large language model with 70 billion parameters	**System message:** Ты утверждаешь решение, принятые другими агентами. **Human message:** Твоя задача - принять окончательное решение по тому, используется ли слово {word} в контексте социальных сетей (имеются в виду элементы интерфейса или действия/концепты, относящиеся к онлайн-коммуникации) или в общем смысле. Твит: "{tweet}" Решение, принятое первым агентом: {agent1_decision} Решение, принятое вторым агентом: {agent2_decision} Проанализируй все доводы, реши, какие кажутся наиболее убедительными, и прими окончательное решение. Отвечай "Социальные сети", если слово используется в контексте социальных сетей. Если слово используется в общем контексте - ставь "Общий смысл". Для ответа используй только эти фразы, не пиши больше ничего.

Table A2. Hyperparameters used for RuBERT fine-tuning

Parameter	**Value**
Optimizer	AdamW
Learning rate	2e-5
Batch size	8
Max length of tokens	128

Epochs	5
Train/test split	80/20

Table A3. Results of RuBERT fine-tuning on a validation dataset

	precision	recall	f1-score	support
0	0.919	0.813	0.863	321
1	0.884	0.952	0.917	479
accuracy			0.896	800
macro avg	0.901	0.883	0.890	800
weighted avg	0.898	0.896	**0.895**	800

Table A4. Distribution of verbs adaptation types across corpus

type	**subtype**	**unique lemmas**	**lemmas %**	**occurrences**	**occurrences %**	**average freq. by lemma**	**occurrences % per type**
borrowed	**transcribed**	47	48.45	6617	29.94	140.79	55.39
	transliterated	27	27.85	5621	25.44	208.19	
translated	**calque**	13	13.4	4401	19.92	338.54	44.61
	extension	9	9.28	5456	24.69	606.22	
	Total	96		22,095			

Figure A1. Proportions for different adaptation types usage per English word over years

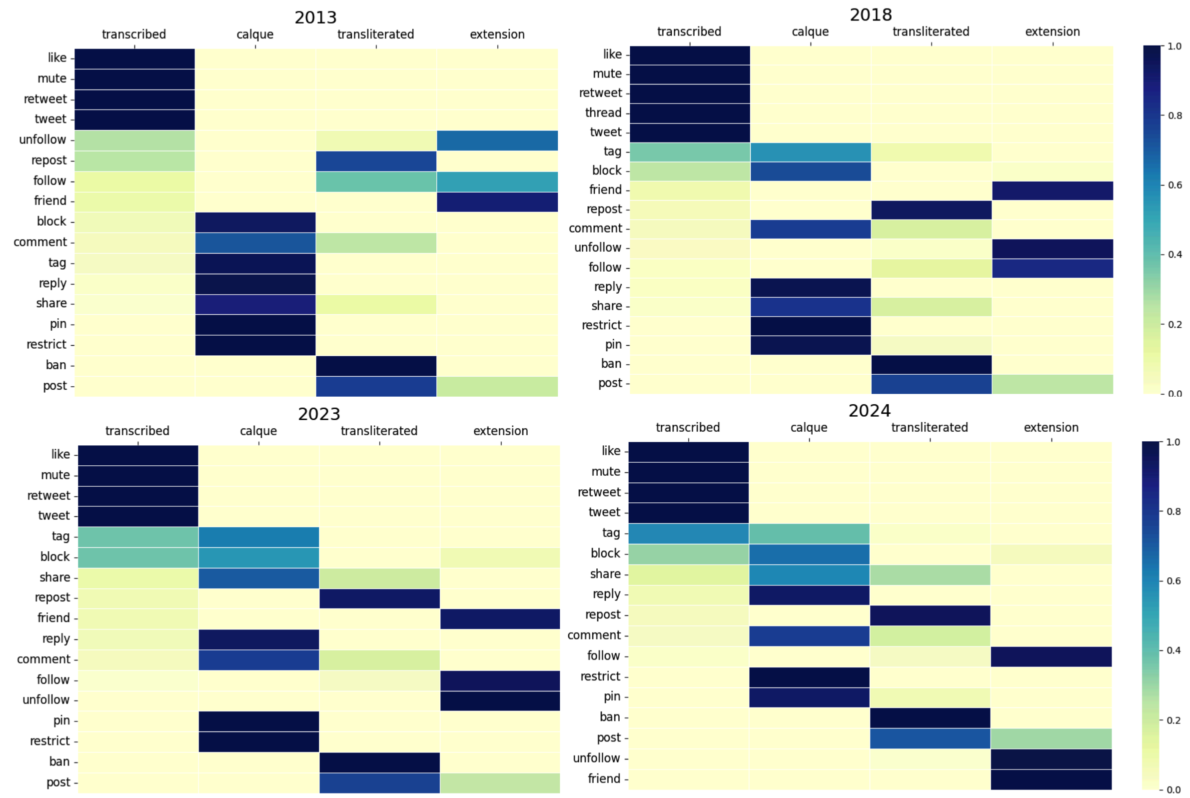

Table A5. Percentage and token frequency of Russian verb forms in different aspects and number and percentage of their unique lemma types

aspect	verb forms tokens	vf tokens %	lemmas types	lemma types %
IMPF	5609	43,81	41	33,61
PERF	7194	56,19	81	66,39
Total	12803			

Figure A2. Trends for different adaptation types usage per English word

Table A6. Distribution of affixes used for perfective lemmas creation counted across the lemmas with the largest number of tokens

Affix	Count	% count
за-	8	44.44
-ну-	6	33,33
рас-	2	11,11
от-	1	5,56
про-	1	5,56

Table A7. English words with all the Russian lemmas denoting them that were found in the dataset, their count and the count of the unique affix combinations they exhibit.

English word	All unique Russian lemmas	Token count (word forms)	Number of unique lemmas	Affix combinations	Number of unique affix combs.	Number of unique inflections
like	лайкнуть, лайкать, отлайкать, полайкать, пролайкать, залайкать, налайкать, перелайкать	2267	8	_нуть, _ть, от_ть, по_ть, про_ть, за_ть, на_ть, пере_ть	8	56
comment	коментить, комментить, прокомментить, откомментить, откоментить, покомментить, прокоментить, комментнуть	200	8	_ть, про_ть, от_ть, по_ть, _нуть	5	24
unfollow	отфолловливать, анфолловить, анфолловнуть, отфоловить, анфоловить, расфолловить, расфоловить, отфолловить, анфоловнуть	327	9	от_ивать, _ть, _нуть, от_ть, рас_ть	5	29
retweet	ретвитнуть, ретвитить, ретвитать, поретвитить, заретвиттить, ретвиттить, ритвитнуть, ретвиттнуть, отретвитить	2113	9	_нуть, _ть, по_ть, за_ть, от_ть	5	41
repost	перепостить, репостнуть, репостить, зарепостить	567	4	пере_ть, _нуть, _ть, за_ть	4	25
ban	забанить, банить, перебанить, забаннить, побанить	1886	5	за_ть, _ть, пере_ть, по_ть	4	35
post	постить, запостить, постнуть, напостить	2056	4	по_ть, за_ть, по_нуть, на_ть	4	29
share	шерить, расшарить, шарить, пошарить, шэрить, пошерить, зашарить, расшерить, зашерить	152	9	_ть, рас_ть, по_ть, за_ть	4	28
friend	зафрендиться, френдиться, френдить, зафрендить	35	4	за_иться, _иться, _ть, за_ть	4	9
mute	мьютнуть, замьютить, мьютить, замьютиться	190	4	_нуть, за_ть, _ть, за_иться	4	19
block	заблокать, заблочить, блокать, блокнуть, блочить, блокирнуть	400	6	за_ть, _ть, _нуть	3	36

report	репортить, зарепортить, репортнуть	12	3	_ть, за_ть, _нуть	3	9
follow	фолловить, зафолловить, зафоловить, фоловить, фолловнуть, фоловнуть	1068	6	_ть, за_ть, _нуть	3	39
reply	реплаить, реплайнуть, отреплаить	69	3	_ть, _нуть, от_ть	3	11
tweet	твитить, твитнуть, затвитить, твиттить	827	4	_ть, _нуть, за_ть	3	23
tag	тэгать, тегнуть, тегать, тэгнуть, тагать, тагнуть, затэгать	132	7	_ть, _нуть, за_ть	3	29
pin	пинить, запинить	2	2	_ть, за_ть	2	2
thread	тредить	3	1	_ть	1	1

Figure A4. Trends for the most popular concepts across the corpus (token frequencies of the Russian lemmas are normalized by the total number of verb counts per year)

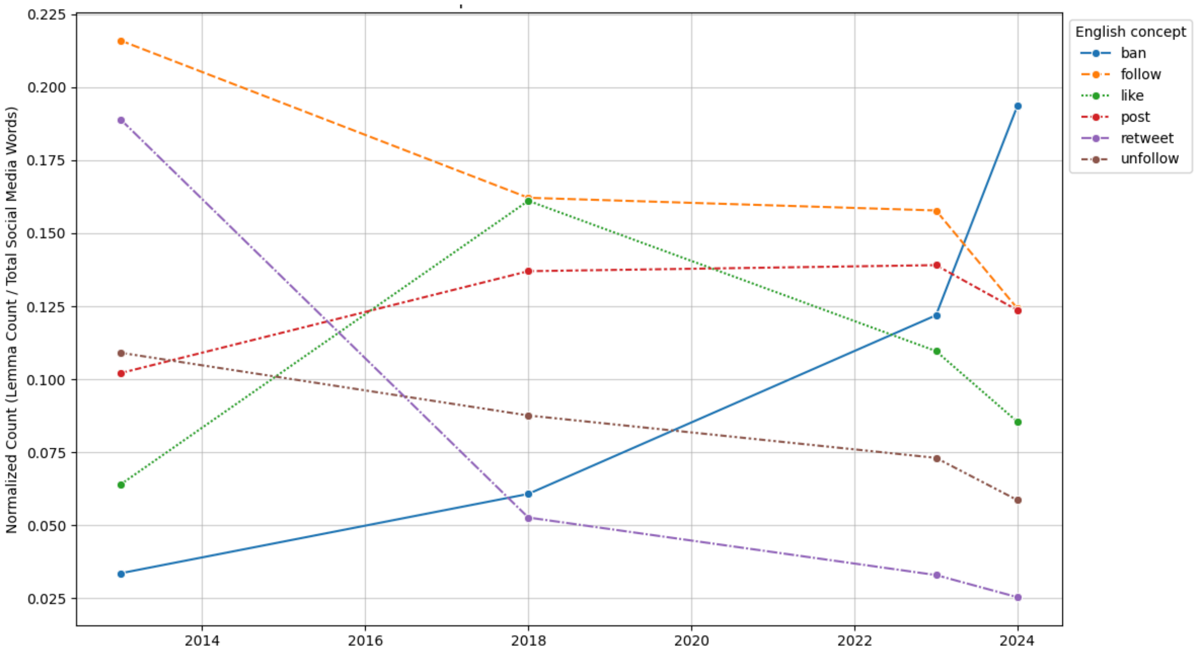

Figure A5. Trends for different adaptation types (lexicalized verbs vs. light verb constructions) usage per English word

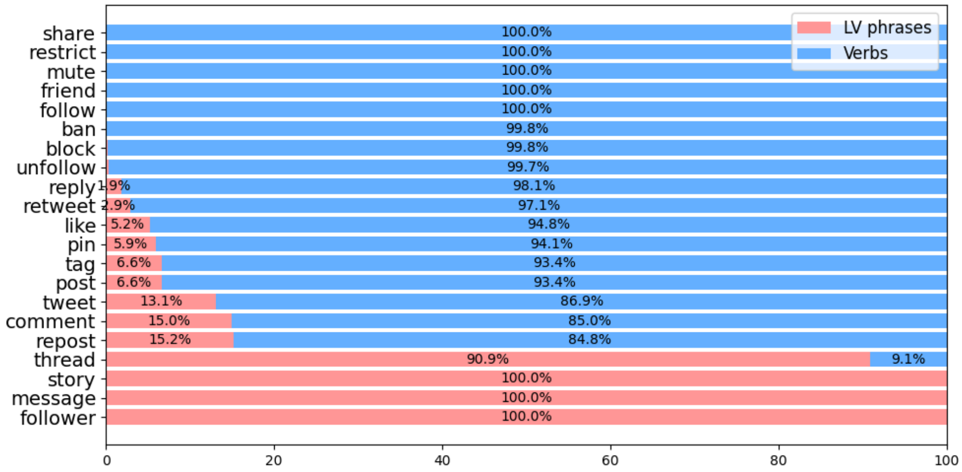

Table A8. English words with all the Russian lemmas denoting them that were found in the dataset, their count and the count of the unique affix combinations they exhibit.

English word	Russian noun (lemma)	All the unique light verbs (lemmas)	Top LV	Total frequency of constr.	Unique constructions
ban	бан	кидать, кинуть	кидать	3	2
block	блок	ставить, сделать, поставить	ставить	3	3
comment	комент	писать, написать, оставлять, оставить, отправлять	писать	20	5
	коммент	писать, оставить, оставлять, написать, выложить, делать, отправить, постить	написать	51	8
	комментарий	написать, писать, оставить, дать, добавить, отправить, оставлять, отправлять, делать, поставить, навалить	оставить	80	11
follow	подписка	оставить	оставить	1	1
follower	фолловер	добавляться, написать	добавляться	3	2
like	like	поставить, ставить	поставить	4	2
	лайк	ставить, оставить, поставить, наставить, нажимать	ставить	119	5
	лойс	поставить	поставить	2	1
message	сообщение	писать, отправить, отправлять, написать, выкладывать, публиковать	написать	65	6
pin	закреп	выложить, оформить, сделать,	делать	6	4

		делать			
post	пост	написать, делать, опубликовать, писать, сделать, разместить, публиковать, выложить, выкладывать, постить, кидать, пилить, отправить, давать, запилить, выкатить, оставить, создавать	писать	189	18
post	публикация	делать, разместить	делать	3	2
reply	ответ	писать, опубликовать, написать, отправить, нажать	писать	12	5
reply	реплай	написать, писать, оставлять, сделать, оставить	написать	22	5
repost	репост	сделать, делать, отправить, ставить, нажать	сделать	91	5
repost	рт	делать, сделать, нажимать, кидать	делать	10	4
retweet	ретвит	делать, сделать, ставить, опубликовать, нажимать	сделать	63	5
story	история	публиковать, постить, кидать, делать, опубликовать	публиковать	8	5
	стори	выложить, выкладывать	выложить	6	2
	сториз	выложить, пилить, выкладывать, опубликовать, постить, сделать	выкладывать	12	6
	сторис	выкладывать, делать, сделать, выложить, постить, запостить, запилить, записывать, записать, кидать, пилить, опубликовать	выкладывать	35	12
tag	отметка	писать, добавить	писать	2	2
	тег	ставить, создать, поставить, написать, писать	ставить	13	5
	тэг	ставить, поставить, сделать, создать, писать	ставить	8	5
thread	ветка	писать	писать	1	1
	тред	сделать, писать, опубликовать, написать, запилить, давать, создать, выложить	сделать	29	8
tweet	твит	писать, написать, делать, постить, публиковать, опубликовать, запостить, выложить, кинуть, сделать, отправить, выкладывать, дать	писать	116	13
	твитт	писать, выкладывать, выложить, постить, написать	писать	8	5
unfollow	анфоллоу	сделать, делать	сделать	5	2

	анфолоу	нажимать	нажимать	1	1