Russian Verbal Affixation in Mental Lexicon: Priming Study and Its Online Replication with True and Stem-Modified Relative Prime Verbs

Chuprina A. O.
NRU HSE, Moscow
a.o.chuprina@gmail.com

Abstract
While suffixed and prefixed words share common lexical features with their base word in the mental lexicon, the two derivational processes have their own properties. Whether their differences are reflected in the mental storage of the group of relative words or not is one of the topical psycholinguistic questions. My experimental results indicate that memory representations of the derivatives differ: while between the stem and the suffixed relative, the relationship is closer and based on transparency of the derived meaning, the relationship between the stem and the prefixed derivative is rather formal. The results also signal that the decompositional route is not a preferred strategy in complex verb processing for a Russian speaker. I base this conclusion on the results of two in-person experiments and their online replicas. Additionally, the data suggest that lexical organization is modified through the aspectual information of family members. These findings need to be taken into account in future studies, both within psycholinguistic and computational fields, using verbal material of the Russian language.

Keywords: Russian verbal affixation; lexical access; mental lexicon; speed-accuracy trade-off; online experimentation

DOI: 10.28995/2075-7182-2022-21-106-113

1 Background
My study investigates how morphologically related verbs are accessed in the mental lexicon while reading a word without a context. Beyond the theoretical objective, the research deals with the problem of result replication. For this, I use the power analysis on the available data in order to plan an online replication study to support the obtained major effects and critically evaluate the emergent trends. This paper concludes with a diagnostic comment on the data collected through online platforms.

Studies of mental lexicon deal with the question of how various types of information carried by a word affect the storage and access to a word in memory both in comprehension and production. The majority of morphological processing studies provide evidence for some effects of universal importance among languages like a facilitatory effect of word frequency, inhibitory effect of word length or age of acquisition ([9] for a review). However, language specific effects are far from being substantially accounted for. Moreover, once they are, the experimental findings fail to be uniformly explained, especially within the theory of decomposition and its mechanisms of affix stripping [22] and edge-aligned embedded word activation [13], [18]. This not only revives alternative theories of whole-word storage [12] and a dual-route model [2], [7] but invites a revision of results through the lens of the morphological phenomenon under investigation. Theoretical conclusions were primarily drawn upon a mixture of grammatical classes, while later experimental data confirmed that the information incorporated into the memory varies depending on the part of speech of a word. Particularly, in Hebrew a priming effect was observed for verbs but not nouns in the condition when both prime and target shared a stem [11]. Similar discrepancies in lexical storage were found for Russian verbs but not deverbal nouns [25] and Buntu nouns but not verbs [4].

My research explores the storage properties of base verbs in relation to their prefixed and suffixed derivatives in Russian. As the base word undergoes various changes, accompanying either of the processes, these differences may play a role for lexical access to the base word. I study this, using an in-person or offline traditional mode and an online platform for the replication study.
2 Material of Russian Verbs

Suffixation and prefixation differ in formal and semantic changes they bring onto the base [14]. A suffixed verb does not share the inflectional class of the base verb, which results in its distinct inflectional paradigm; and its stress may shift from its positional morpheme in the base. The base verb undergoes phonological transformations (vowel/consonant truncation, insertion, and alternation) for the morpheme boundary of the neighboring stem and suffix to comply with the phonological rules of Russian. However, the meaning of the related word in suffixal transformations of two most common types that I use in this study is uniformly predictable: the suffix -nu- signals a single act of the base verb-denoted action, and the suffix -va- indicates the process of imperfectivation.

The prefix, on the other hand, does not imply as many changes of the base verb: the two relative verbs share the verbal class and most often the stress pattern. On the contrary, the related verb is not always semantically transparent outside the context (затолкать ‘push/push into’ in the meaning of ‘pushing’, ‘forcefully hitting somebody in a crowd’ or ‘starting to push’). In addition, the aspect of the base verb changes only when an imperfective base verb undergoes the prefixation, but not the perfective one.

Additionally, an exploratory corpus investigation showed a difference in use of these two derived forms. The dominant form of the prefixed verb surfaced as that of the participle (1:2, for each finite form there are two non-finite forms of a prefixed verb). On the contrary, a suffixed verb primarily appears as a fully inflected form, which is comparable to the syntactic behaviour of the base verb (1:3). Overall, this suggests a functional difference between the two relative verbs.

The study of verbs and their derivatives in Russian unavoidably includes aspectual information into the research frame. Prior studies of verbs [23] and theoretical discussion of the aspectual pairs ([27] among many others) suggests a closer mental organization between a verb and its imperfective -va-derivative, while the connection between the base and its prefixed relative verb might depend on a particular prefix or overall might be disconnected from the base, acquiring its own semantics with a distant relation to the base [16]. In the first case, the imperfective verbs will provide a processing advantage to their bases, resulting in shorter reaction times upon reading the base. In the second case, prefixed verbs will give very little facilitation to their base during lexical access and show a longer decision time to the base.

To sum up, using a group of verbs of Russian I compare the properties of lexical access to them after a brief exposure to their related affixed verbs. The main prediction is that the faster the access is to the base, the closer the storage between the base and its derivative. If the suffixed derivative is closer, the mental connection between the relative verbs is based on the clearly related semantics between the two, easily overriding formal changes of the derivational process. If the prefixed derivative is closer, the mental organization depends more on the formal resemblance between the two verbs. The continuation of the study, using the copies of the derivatives as primes but with an orthographically transformed stem, will test the presence of the decompositional strategies that can be used during access to the base verb. In particular, the affix stripping and edge-aligned embedded word activation operations will put a connection between the base and its prefixed derivative into an advantageous position during access. While none of these particular operations could account for the type of the suffixal connection in Russian, if the access is still faster from the quasisuffixed verb, it will point at a somewhat two-stage decomposition that strips a verbal suffix -т’ and spots -nu suffix simultaneously. If decomposition is not a preferred strategy for accessing a Russian complex verb, I cannot expect any facilitation from neither of the relative verbs.

3 Data and Procedure

I use the experimental procedure of the lexical decision task with masked morphological priming [10], see the Data and Procedure section for the details. This experimental procedure does not demand a substantial expenditure of the resources on the part of a researcher and has validated itself on many decades of research on different languages. It is not surprising that online data collection platforms now allow researchers to use the available code or create it from scratch and run their priming experiments [1]. I ran two offline experiments and their two online replicas. In Experiment 1 ‘Real Affixed Verbs’ primes were suffixed and prefixed derivatives of the base verbs as targets. Experiment 2 ‘Quasistem Verbs’ used the same target verbs, but the primes were nonwords obtained by replacing a letter within a stem and an intact affix of the real verb used in Experiment 1 (see Table 1).
For the target material I selected 39 target base verbs. The length of the stimulus words was five to nine letters, their own frequency was in the range from zero to 70 ipm (occurrences of a word per million in the main corpus of the Russian National Corpus). The frequency was taken from the Frequency Dictionary of the Modern Russian Language, edited by O. N. Lyashevskaya and S. A. Sharov [20]. A group of unrelated verbs was chosen to control the results for related prime conditions. Afterwards, an additional group of 13 fillers was added to equal the lists in related and unrelated conditions overall. To these 52 targets, I added the same number of nonword targets, preceded by real word primes in Experiment 1 and nonword primes in Experiment 2. It should be noted that nonword targets in Experiment 2 appeared in the related and unrelated conditions, i.e. preceded by relative and unrelated nonword primes [6]. All nonword primes were constructed, following a natural phonetic structure of the Russian language.

The experiment consisted of performing a lexical decision task with visual masked priming. Each experimental attempt by a subject consisted of seeing a screen with an asterisk, serving as a fixation point (varied between 1000ms and 1500ms); a 500ms mask, the length of which corresponded to the average length of the target words (seven bars, #); a 60 ms prime and a following 20ms mask. The function of grids or masks is to avoid the visual trail of words appearing on the screen. Finally, a 500ms target word screen appeared between two blank screens of 30ms and 2000ms correspondingly. To respond, a subject pressed the key – either ‘left Ctrl’ or ARROW RIGHT, which were located in the bottom row of the keyboard. The participant's task was to decide as quickly as possible whether they see a real word, or whether the word does not exist by pressing the corresponding key. The speed and correctness of the answers were measured.

The experiment was carried out on a DELL personal computer using the OpenSesame program [21]. The overall duration of the procedure amounted to seven minutes.

The sample consists of 39 participants in Experiment 1 and in Experiment 2 of 25 participants. In online replication studies the samples consist of 56 and 41 in Experiment 1 and 2 correspondingly.

All speakers speak Russian as their native language. The age range of the entire sample is between 18 and 59 years of age with female participants constituting two thirds of the whole sample. All participants completed the task in full and were not aware of the objectives of the study except for the general psycholinguistic research into the process of reading.

4 Results

I analyzed the reaction of the subjects according to two parameters: the accuracy of the choice "word or nonword" and the speed of the response. Based on the general distribution of the respondents' answers, I determined the threshold of 80% accuracy, below which it could be uncertain that the participant in the experiment gave thoughtful answers. As a result, the data of nine participants in Experiment 1 and four participants in Experiment 2 were not included in the statistical analysis. To describe the data, I used the generalized additive model as in package mgcv [26] in the R programming environment for statistical data processing [24].

I now present the data of both experiments with real (Experiment 1) and quasistem primes (Experiment 2) for immediate comparison of the effects.
In Experiment 1 the response accuracy showed a subject random effect (Chi.sq=132.1, p<0.01) as well as main effects of target frequency (Est.=0.7, SE=0.09, p<0.01) and target aspect (Est.=-1.02, SE=0.23, p<0.01), imperfective group having an advantage over the perfective one. An effect of prime type was not observed. In Experiment 2 the analysis showed the significant random effects of the subject (Chi.sq=28.1, p<0.01) and target (Chi.sq=31.6, p<0.01) as well as main effects of target’s frequency (Est.=0.4, SE=0.15, p=0.002), a positive effect of length of the target (Est.=0.86, SE=0.33, p=0.01) and the negative one of the prime (Est.=-0.7, SE=0.3, p=0.01). From the group of related verbs, the prefixed prime showed a positive effect on accuracy in comparison to unrelated controls (Est.=0.98, SE=0.4, p=0.02).

Before analysing the reaction time data, I excluded the data points for incorrect answers and extreme values greater than 2500ms, following an established data curation procedure [4], [5], [19]. To this, I lost 10% of each experiment’s data sample. Also, a quadratic transformation of the dependent variable was performed, guided by the boxcox plot suggestion [3].

The following analysis revealed the random effect of the subject (F=24.16, p<0.01) in Experiment 1 and subject (F=20.69, p<0.01) and the target verb factors (F=0.49, p=0.05) in Experiment 2. Beyond main effects of frequency (Est.=−4.8−0.05, SE=9.1−06, p<0.01) and aspect, perfective targets recognized more slowly (Est.1.4−03, SE=3.4−04, p<0.01), in Experiment 1, the morphological type of the prime verb also showed an explanatory power not only in comparison to the unrelated verb condition (prefix facilitation: Est.=−1.02−03, SE=3.2−04, p<0.01; suffix facilitation: Est.=−1.8−03, SE=3.2−04, p<0.01) but also among themselves, a suffixed verb having an advantage over a prefixed derivative (Est.7.2−04, SE=3.3−04, p=0.03). As for Experiment 2, the effects of frequency (Est.8.5−04, SE=1.5−04, p<0.01) and aspect (Est.1.5−03, SE=3.9−04, p<0.01) were the only ones to account for the variance in the data sample.

There are various methods to calculate the plausibility of the null hypothesis, depending on the statistical model and the type of data. For example, power analysis, using simulations, can be applied for generalised linear effect models, while Bayesian inference – for more nuanced complex modelling. To fit into the bigger picture of morphological processing research, where the data are mainly analysed with linear models, I chose the power analysis with simulations. The effect of interest after running the offline version was the suffix ‘advantage over prefix’. The power calculation for the model to detect the suffix effect of this model was 60% which is good but not sufficient for the certainty in theoretical conclusions. With the help of extend method from the simr package [15] I identified the necessary subject sample size of 40 to detect the effect with about 80% confidence.

To program the replication of Experiment 1 and 2, I adjusted the masked priming experiment template available at PCibex, an online experiment builder [28], and to recruit the participants I used Toloka, a Yandex crowdsourcing platform [29]. Due to technical differences between programming procedures for the original study and its replication, some issues arised. One of them is that the duration of the fixation point was always 1500ms in comparison to its varied 1100−1500ms timing in the offline experiments. To see whether this factor should be corrected for, I included it into the initial offline model and saw that it does not affect the data variation. The other issue was that the online response could only be given upon the disappearance of the target off the screen, i.e. there was no chance for a participant to react with a key press within the target duration of 500ms. This could have been a problem if the RTs’ distribution and its central tendency measures suggested an anomaly in the online data sample in comparison to the offline data (offline RTs: mean (695.2ms), sd (237.1), median (643); online RTs: mean (920.2ms), sd (364), median (835)). This was not the case. The online RTs data were overall slower than in the offline sample but both samples were distributed identically.

As before the data for subjects that gave too many false responses or did not finish an experiment were extracted from the further analysis, making a 45-subject sample in Experiment 1 and a 34-subject sample in Experiment 2. In online Experiment 1 the response accuracy revealed the subject random effect (Chi.sq=207.8, p<0.01) as well as main effect of frequency (Est.=0.29, SE=0.08, p<0.01) without the aspect effect as it was observed in the offline version. In Experiment 2 the significant random effect was the one of the subject (Chi.sq=55.02, p<0.01). The main effects in online Experiment 2 were of target verb frequency (Est.=0.57, SE=0.1, p<0.01), length of the target (Est.=0.59, SE=0.24, p<0.01) and of the prime (Est.−0.44, SE=0.22, p=0.05) in the same direction as in the offline mode. As before, the prefixed prime showed a significant positive effect in comparison to unrelated controls (Est.0.57,
SE=0.31, p=0.07) but this time, including the contrast to the suffixed prime, suffixed primes causing more mistakes (Est.=-0.8, SE=0.3, p=0.02).

After cutting incorrect responses and extremely high values (14% of the data sample in online Experiment 1 and 13% in online Experiment 2) I analysed the RT data. The analysis revealed the random effect of the subject (F=20.64, p<0.01) and target (F=0.8, p<0.01) in Experiment 1. The same factors were observed in Experiment 2. Among the main effects in Experiment 1, there emerged an effect of frequency (Est.=-0.0005, SE=0.0001, p<0.01) and both relative primes in comparison to unrelated controls but not among themselves as it was in the offline sample (prefix facilitation: Est.=-0.001, SE=0.0003, p<0.01; suffix facilitation: Est.=-0.0008, SE=0.0003, p=0.003). As for Experiment 2, the effect of frequency (Est.=-0.0005, SE=0.0002, p<0.01) was the only robustly significant effect while the others including the related primes emerged at the level of tendency.

After cutting incorrect responses and extremely high values (14% of the data sample in online Experiment 1 and 13% in online Experiment 2) I analysed the RT data. The analysis revealed the random effect of the subject (F=20.64, p<0.01) and target (F=0.8, p<0.01) in Experiment 1. The same factors were observed in Experiment 2. Among the main effects in Experiment 1, there emerged an effect of frequency (Est.=-0.0005, SE=0.0001, p<0.01) and both relative primes in comparison to unrelated controls but not among themselves as it was in the offline sample (prefix facilitation: Est.=-0.001, SE=0.0003, p<0.01; suffix facilitation: Est.=-0.0008, SE=0.0003, p=0.003). As for Experiment 2, the effect of frequency (Est.=-0.0005, SE=0.0002, p<0.01) was the only robustly significant effect while the others including the related primes emerged at the level of tendency.

To evaluate the main finding of suffix’ greater facilitation over prefixes in the offline mode for the real verbs’ pair and the effects at tendency levels in the online mode after the quasistem primes, I combined offline and online datasets into one and included the mode of the experiment as an independent variable. In Experiment 1 with real primes the effect of the mode of the experiment was not observed in the accuracy data. However, the speed of online responses was significantly slower than the offline ones (Est.=4.6−03, SE=6.8−04, p<0.01). The effects of frequency and target aspect were preserved as well as the ones of prime type, suffix condition being faster when compared to the unrelated primes (Est.=-1.8−03, SE=3.1−04, p<0.01) and their prefixed counterparts (Est.=-7.1−04, SE=3.2−04, p=0.03). In addition, the interaction between the mode and prime type showed a significant slow down for suffixed primes online (in the comparison to the unrelated controls: Est.=-9.8−04, SE=4.1−04, p=0.02; and to the prefixed primes: Est.=-9.5−04, SE=4.1−04, p=0.02). However, calculating the difference between the two slopes and accounting for the standard errors, the suffix effect was still facilitatory in contrast to both the unrelated and the prefixed primes.

The combined data analysis for accuracy in Experiment 2 did not show data variation for the effect of the experimental mode and the previously observed effects of frequency and lengths of target and prime were robust as well as the beneficial effect of prefixes over the unrelated condition (Est.=0.9, SE=0.4, p=0.02). The reaction times data also showed a significant slow down in the online sample (Est.=3.6−03, SE=7.6−04, p<0.01) as well as the main effects of frequency (Est.=-5.4−05, SE=1.4−05, p<0.01) and target aspect (Est.=-1.7−03, SE=5.1−04, p=0.01). An interaction surfaced between the mode and the prime time which indicated a greater facilitation from prefixed primes over unrelated ones online in comparison to the identical case in the offline experiment (Est.=-9.1−04, SE=4.5−04, p=0.3).

Finally, the error analysis of data on nonword prime-target pairs in the offline Experiment 2 showed the significance of the same random factors as in the analysis of the material of real words. The morphological type of the verb was not found to be significant in explaining either the variability in the correctness data or the reaction times.
5 Discussion

Morphologically related words are stored especially closely in the memory of speakers of different languages. My study assessed whether for Russian speakers a particular type of the morphological relation is reflected in the mental lexicon. My study on derived suffixed and prefixed verbs shows that there are psychological differences in the organization of storage of and access to the base verb, channeled through either of the derivatives, a result similar to the findings in Korean [19].

The robust observed effects showed an importance of the stem morpheme for the transference of lexical information from an affixed word to its base: once the lexical relatedness was cancelled by a single-letter modification within the stem of the derivative, the advantage from the related primes over unrelated primes was cancelled as well with no regard to the intact affixal morpheme. Instead, we could observe purely orthographic effects, ones of target’s and prime’s lengths, to emerge for accuracy of response. The fact that these effects were accompanied by the prefix effect in the quasi-affixed prime data could also point to the orthographic nature of the base-prefixed derivative connection. An approximation of the decompositional route that a Russian verb may go through is an affix stripping or more specifically, a prefix-stripping mechanism. However, the cognitive nature of this mechanism deserves a closer look as the prefixal advantage was observed only for accuracy measure.

The speed of response results show that suffixed derivatives are better facilitators of the base verb’s recognition. The presence of a real suffix on the root that deviated from the shared stem by a single letter did not give a processing advantage over an incompletely unrelated letter string. Together it shows that the psychological representation of the connection between the base and its suffixed derivative is localized inside the shared root of the pair. This, in contrast to the prefix’s orthographic relation, indicates the lexical nature of the suffix relation to its base.

The emergence of the facilitatory suffix effect for the reaction time in the experiment with real primes and of the beneficial prefix effect in the accuracy results in the experiment with nonword primes suggest that multiple memory processes mediate morphological connections of a base word with its derivatives. This is in line with the studies on speed-accuracy trade-off [17]. This also brings about the next observation about the verb’s aspect factor, which emerged in the offline data samples and was more robust for the real verb pairs. I cannot extrapolate this to a single verb recognition due to the methodological two-word nature of the study, but my results signal more common information stored in memory between the imperfective verb and its perfective derivative than inside the pair of the reverse relationship. This result along with a previously found aspect effect in sign language [8], deserves further investigation for the benefit of both theoretical and cognitive linguistic fields.

As for the available modes for experimentation, the researcher should be aware of the discrepancies that might arise. Although the online mode is an undeniably valid tool for the data collection and a valuable platform for replication studies, one should take care of the smaller volatile effects. Due to the longer speed of response in the offline mode, one can possibly detect another processing procedure than the one hypothesized. Specific to my case, this is the effect of the suffix in relation to prefix, showing discontinuity in the online mode of experimentation. Since the extra time, taken online before an actual button press, allows for more prime-target information becoming available, I can tentatively conclude that the lexical properties being incorporated at later time points might also differ for either of the morphological processes.

Altogether, my study not only confirms that mental lexicon reflects various means of organization of words but also is one of the few that demonstrates that even within a homogeneous group of a single category of verbs one can expect a substantial variation in the mental organization.

Acknowledgements

I am deeply grateful to my PhD supervisor professor Nataliia Slioussar at HSE (Moscow, Russia) for introducing me to the research of complex word storage in the mental lexicon, for suggesting the research method, and for editing the initial verb lists.
References


[20] Lyashevskaya O. N., Sharov S. A. (2009), Frequency dictionary of modern Russian language (based on National corpus of Russian) [Chastotnyj slovar' sovremennogo russkogo yazyka (na materialah Nacional'nogo korpusa russkogo yazyka], Moscow; Azbukhovnikh.


[28] Zehr, J., & Schwarz, F. PennController for Internet Based Experiments (IBEX), 2018 https://doi.org/10.17605/OSF.IO/MD832

[29] https://toloka.yandex.com/