MORPHOLOGICALLY MEDIATED EFFECT OF SYNTACTIC DIVERSITY

Anastasia Chuprina[1,2], Nicholas Lester[3], Natalia Slioussar[2,4]

1 – Institute of Slavic Studies, University of Graz
2 – School of Linguistics, HSE, Moscow
3 – Department of Comparative Linguistics, University of Zürich
4 – Department of the Problems of Convergence in Natural Sciences and Humanities, St.Petersburg State University
a.o.chuprina@gmail.com

ABSTRACT

In the mental lexicon words are connected to each other through various paths. We explore how a word's representation might be accessed, depending on its syntactic properties and shared formal properties with other members of a morphological family. Morphological families of verbs in Russian are primarily related through processes of prefixation and suffixation. Based on the changes that these two processes introduce to the derivational paradigm and the base word's syntactic distribution, we expect to observe a resultant adjustment of the lexical access to the family's stem. We used the data obtained in two priming studies of morphologically related verbs in Russian. The results account for a facilitative effect of the syntactic variability of a stem verb observed after pre-activation of the suffixed relative, while that of the prefixed verb showed an inhibitory influence.

Keywords: morphological relatives, syntactic diversity, lexical access

1. BACKGROUND

Lexical access to a word is known to be influenced by a word's relation to other words. The words share information through different channels – inflectional paradigms [9], derivational family size [11], semantic neighbourhood [4]. Derivational families in morphologically rich languages are the material to investigate an interaction between these types of information.

Previously, it was found that an activation of a prime accelerates the lexical access to its syntactically coherent target [6, 12]. More importantly for the present research, syntactic properties of a word influence the word's recognition in isolation [7]. In our turn, we explore whether shared syntactic features, sheltered behind common formal properties, play a role in the lexical access to the stem word of the morphological family.

Russian language builds new words, making use of both prefixation and suffixation. Depending on these derivational processes, the lexical properties of the derived words differ. In the verbal domain, prefixation usually preserves all inflectional properties of a verb except for the aspect and results in relatively unpredictable semantic changes. Suffixation, however, is semantically more predictable, but always changes the verb's inflectional class. Importantly, in case of the prefixation, the resultant verb may change its syntactic distribution in comparison to the stem. For example, a prefixed verb, allows for a combination with a prepositional phrase (*лететь-взлететь на что-то,* 'fly somewhere-fly up onto something').

Additionally, a preliminary corpus study showed a difference in use of these two derived forms. The dominant form of the prefixed verb is 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, for a non-finite form of a suffixed verb there are three finite forms). Overall, this suggests a functional difference between the two relative words.

Our observation also corresponds to a previous finding of the study of the inflectionally prolific language of Estonian [9]. The authors applied a new measure of a number of actually occurring paradigm members for a given word, available in the 15-million token Balanced Corpus of Estonian. The inflectional paradigm size counts

were based on inflected forms in real use rather than on the number of forms that an Estonian noun paradigm has in principle. This diminishes the role of the potential volume of the paradigm that a language is capable of but elevates the language use as a function of an actual lexical volume that is produced and processed by language speakers.

Our data adds to this research, linking the number of actual forms to their dominant function. Together, those properties may be of value for the tissue of morphological relatedness between the words.

Based on the above, we expect that suffixed and prefixed words as primes will interact differently with the syntactic information carried by the base form. More precisely, we expect suffixed forms, which syntactically behave in line with the base verb and allow for a similar syntactic variability, to serve as more substantial primes than prefixed verbs with limited syntactic potential.

2. METHODS

Pre-activation of morphologically relative representations affects the speed of the base word recognition. We are interested in whether the observed morphological effects will interact with the target's own syntactic distributions.

We employ a priming paradigm in which the targets are the base forms and the primes are target's morphological relatives. The syntactic distribution of a base form is defined as its frequency distribution across arc labels in a hand-annotated dependency parse of naturally occurring Russian text.

2.1. Measuring syntactic diversity

Based on the overall entropy of a word (see Equation 1), we measure the syntactic information carried by the verbs' syntactic distributions. We do this, using the conditional entropy (see Equation 2) where D is syntactic dependencies that a word is a part of, and L is lexical associates of a word. This measure allows us to isolate syntactic information from the information carried by specific lexical items that are syntactically bound to the target (i.e. collocational information). Taking the dependency grammar inventory of the syntactic functions, we refer to these measures as "modifier diversity", "head diversity" and "total syntactic diversity", where total syntactic diversity is defined as the sum of the head and modifier entropies and the entropy of the choice between the two syntactic functions.

(1)

$$H(x) = -\sum p(x) log p(x)$$

(2)

$$H(D|L) = H(D,L) - H(L)$$

This formulation represents the most economical operationalization of syntactic distributions, while also allowing researchers to explore the interplay of a word's relative status (head vs. modifier) and word order.

For the current research, we computed the frequency distributions of the base form of each verb across its syntactic dependencies in the largest available dependency treebank of Russian SynTagRus, which contains approximately a million words. We then obtained the measure of syntactic distributions of the verbs and added this information to the primed lexical decision dataset. Due to the limitations of the corpus, the syntactic measures were only available for the target verbs of the dataset. Also, we were left with two syntactic measures for our dataset, the total and modifier syntactic diversities. The fact that the verbs in our list appeared incompatible with the role of the head deserves a further exploration. The current state of the annotation for the word entry marks the word as a root, the head of the whole sentence, or a modifier, but not the head of a phrase (e.g. *он увидел, как вода капает с крыши* "he saw, how the water was dripping from the roof", *was dripping* is altogether a modifier of the *saw* and a head of the noun phrase *the water* and the prepositional phrase *from the roof*). To presently correct for this, we residualised the modifier component from the total diversity measure and used the outcome in our analysis as a measure of probability of the verb to be used in the other syntactic function.

Crucially for this research, we expect the participial preference of the prefixed forms to reflect a strong bias towards modifiership, and hence to interact with the syntactic diversity of the target differently from the suffixed verbs' pre-activation.

2.2. Data and Procedure

In our research, we made use of the experimental paradigm of lexical decision with masked priming [5]. The task, performed by a participant, is to read a letter string, a target, which is briefly preceded by another word, called a prime. By pressing a button, participants respond whether the target string is a real word or a non-existent letter sequence. During the experiment, the accuracy and the reaction times are recorded.

We used behavioural data from [3] which is the result of two masked-priming lexical decision experiments on Russian verbs with prime duration of 150ms and 60ms (41 participants). Primes were derived verb forms while targets were their stems. All the verbs appeared in the form of the infinitives.

Originally, the dataset consisted of 37 morphological families of verbs. The derivative prime-verbs were balanced in terms of their token frequency, length and orthographic relatedness to the target measured by the Levenshtein distance [8]. After we had merged the syntactic diversity measures and the experimental data, our final list shortened to 20 imperfective target verbs and their perfective suffixed and prefixed relative forms as primes of the initial dataset.

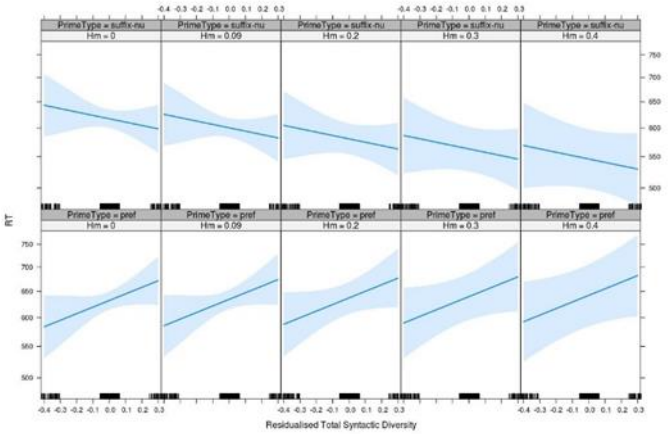
3. RESULTS

A preliminary analysis revealed no difference in the effects for each condition, so we fitted a mixed effect regression model [1] on merged 150ms&60ms RT data, predicting response latencies as a function of several variables: that of the type of morphological relative verb and the syntactic diversity of the target verb. In addition to the fixed effects, we included random effects for participants and the sequential position of each trial in the overall experimental order of the presentation. We discarded 6.1% of all trials as outliers (all latencies falling below or above 2 standard deviations of the mean). We also took the logarithm of the response times as suggested by a Box-Cox power analysis [2]. After that, we included the syntactic measures into the model. Accuracy analysis revealed an expected effect of the target's frequency, the higher the frequency the more correct answers are given. The measures of diversity elicited only a marginally positive effect on the part of the target's function as a modifier at p=0.07. No interactions with the preceding morphological relatives were captured.

The reaction times analysis replicated the original finding of [3]. Both types of relative verbs compared to the unrelated primes had a facilitative effect on the reaction times, to a greater degree in the suffixed group than the prefixed alternative (β=-0.04, SE=.02, p=.02).

The independent effect of the primes was also preserved when we included their interactions with the target's syntactic measures into the model. Importantly, the prime type and syntactic diversity measures showed a two-way interaction. The reaction was significantly accelerated by the modifier diversity of the target, augmented by the preceding suffixed prime activation (β=-0.31, SE=.10, p<.01). On the opposite, the speed of the decision-making was inhibited by the target's residualised total diversity, paired with the prefixed prime pre-activation (β=.20, SE=.10, p=.05) as visible in Figure 1.

Figure 1: Reaction times to the stem verb, based on its Modifier and residualized Total Syntactic Diversity, primed by suffixed (upper panel) and prefixed (lower panel) relative verbs.



To further counteract data sparsity issues, we binned the entropy measures into three groups based on the proportion of observations: low, medium, and high. The analysis reproduced the finding, but additionally, now the effects showed a sign of graduality. The two syntactic measures revealed the interactions at medium and high levels of diversity. The prefixed primes showed a significant inhibition at mid-level of residualised total diversity (β=.1, SE=.05, p=.05). On the opposite, the suffixed primes were significantly facilitative at mid-level of target's modifiership (β=-0.1, SE=0.03, p=.00)

and at high level of the residualised total diversity measure (β=-0.08, SE=.03, p=.01).

4. DISCUSSION

We investigated how syntactic connections of a word might influence its recognition, modified by its relation to the morphological relatives. Our results reveal a two-way interaction between the morphological and syntactic properties of the derivationally related verbs. Essentially, morphosyntactic properties play a role for the lexical access to the stem of the derivational paradigm.

The effect from the suffixed primes is facilitatory both individually and within the interaction with the syntactic properties of the target, while that of the prefixed primes is facilitative on its own, but taken together with the stem's syntactic distributions, is inhibitory. Due to the lack of data points within the dataset, we cannot account for the base verb's full syntactic variability.

The observed interaction between the relative verbs' morphological properties and the target's syntactic variability was characterized by the bidirectionality of the effect. We explain this as a function of the syntactic similarity on the part of the suffixed relative and the diversion of the syntactic behaviour that is introduced by the process of prefixation in relation to the stem's syntactic distribution.

The result also suggests a graded organization of the observed morphosyntactic effect. That means that the more connections a base word builds, the greater is the difficulty to recognize it after the pre-activation of the syntactically conflicting relative. By pre-activating a form with more limited syntactic potential, the base form with a higher syntactic pluripotentiality is recognized slower as a function of the fewer points of syntactic distribution between the two as in [7] for further evidence from English.

However, we cannot attest further to the nature of the observed effect and consider the present outcome as a hint for a more detailed exploration of the interaction between the morphological and syntactic properties of the relative words.

In sum, syntactic connections, which a word participates in, are stored in memory and channelled through the links to the members of the derivational family. We hereby conclude that the storage of a word and access to it are dependent on the morphosyntactic connections, established within a word's family.

REFERENCES

1. Bates, D.M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.
2. Box, G., & Cox, D. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*. Series B (Methodological), 26(2), 211-252.
3. Chuprina, A. (*submitted*). Morphological effects on lexical access to a base word.
4. Feldman, L. B., Marelli, M., Amenta, S. and Milin, P. (2015). Effects of target density on morphological processing at long and short SOAs. *First Quantitative Morphology Meeting*. Belgrade, Serbia. 11-12 July 2015.
5. Forster K. I., Davis C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition. Vol. 10*(4). P. 680-698
6. Goodman, G. O., McClelland, J. L., & Gibbs, R. W. (1981). The role of syntactic context in word recognition. *Memory & Cognition*, *9*(6), 580-586.
7. Lester, N. A., Feldman, L. B. & Moscoso del Prado Martín, F. (2017). You can take a noun out of syntax…: Syntactic similarity effects in lexical priming. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. p. 2537-2542.
8. Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*. 10 (8): 707–710.
9. Lõo, K., Järvikivi, J., Baayen, R.H. (*submitted*). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing.
10. Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227-234.
11. Moscoso del Prado Martin, F, Bertram, R., Häikiö, T., Schreuder, R., Baayen, R. H. (2004). Morphological family size in a morphologically rich language: the case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory, and Cognition. Vol 30*(6), 1271-1278.
12. Röder, B., Demuth, L., Streb J. & Rösler, F. (2003) Semantic and morpho-syntactic priming in auditory word recognition in congenitally blind adults. *Language and Cognitive Processes, 18*:1, 1-20.