Slovo. Journal of Slavic Languages, Literatures and Cultures ISSN 2001–7395
No. 63, 2023, pp. 23–39

From machine learning to classroom learning: mobile vowels and the Russian preposition *v* ‘in(to)’

Tore Nesset and Kevin Xavier
UiT The Arctic University of Norway
Tore.nesset@uit.no, kxa000@post.uit.no

Abstract

The present study reports on a machine learning experiment concerning mobile vowels in the Russian preposition *v* ‘in(to)’. It is shown that a neural network is able to predict mobile vowels in 97.4% of the cases in our dataset, and a decision tree is used to extract a set of three rules that a language learner can use to achieve nearly the same level of accuracy. We argue that these rules are valuable from the perspective of language pedagogy, but that some adjustments are necessary in order to make the rules simpler and more precise. Our study lends support to earlier analyses which emphasize the capacity of mobile vowels to prevent sequences of identical segments. We advance the Word Onset Hierarchy, which enables us to evaluate the relative importance of phonological features for mobile vowels and highlights the gradient and asymmetric nature of mobile vowels. It is suggested that machine learning represents a valuable tool for language pedagogy, not only for mobile vowels, but also for other areas of Russian grammar that are challenging for students of Russian as a foreign language.

1. Introduction

Students of Russian as a foreign language discover very early that some words have forms with vowels that disappear in certain environments. Thus, in (1) the pronoun meaning ‘whole, all’ has a vowel in the stem (*ves’*), while in (2) this vowel is not present (*vsem*). The preposition meaning ‘in(to)’, on the other hand, has a vowel in (2), where it is realized as *vo*. This vowel is not found in (1), where the preposition is realized as *v*:[1]

[1] These examples are from the Russian National Corpus (www.ruscorpora.ru).

(1) Ja vljublen **v ves'** mir. (Kataev 1980–1981)
'I am in love with the whole world.'

(2) Roditeli **vo vsem** mire žalujutsja na to, čto detej nevozmožno otorvat' ot komp'jutera. ("Daša" 2004)
'Parents all over the world complain that it is impossible to drag kids away from their computers.'

Such vowel~zero alternations known as "mobile vowels" represent a major challenge for learners of Russian as a foreign language, and many instructors have received the question "what is the rule?" – a far from trivial question to answer.[2]

Mobile vowels are not only challenging for language pedagogy, they also represent one of the classic theoretical problems of Russian phonology. Simplifying somewhat, the theoretical literature addresses the following questions: Are the vowel~zero alternations we can observe the result of epenthesis or deletion (see Bethin 1998: 210–211, Scheer 2011 and 2019: 202–203 for discussion)? Are there one or two underlying vowels that produce vowel-zero alternations (e.g., Iosad 2020)? What is the conditioning environment for the alternations (e.g., Klapper 1993, Gouskova 2012, Gouskova and Becker 2013, Becker and Gouskova 2016, Linzen, Kasayanenko and Gouskova 2013)? Prepositions figure prominently in the theoretical literature about mobile vowels, since they are relevant for the relationship between prepositions and prefixes – a much debated issue in Russian morphosyntax (e.g., Matushansky 2002, Steriopolo 2007, Gribanova 2009, Blumenfeld 2012).

In this article, we consider the conditioning environment of mobile vowels from the point of view of language pedagogy. We concentrate on the preposition *v* 'in(to)' and ask whether a machine learning experiment can help us formulate rules that can be useful in the classroom. Our contribution can be summarized as follows. First, we show that a neural network is able to learn to choose between *v* and *vo*; the model predicts the correct variant in 97.4% of the cases in our dataset of 18,207 preposition-word samples. Second, we develop a decision tree and argue that the model can be translated into rules that may be of pedagogical value. Third, it is suggested that the rules that emerge from the experiment may be made simpler and more precise. This, we argue, suggests that machine learning is a valuable tool in language pedagogy, but cannot replace linguistic analysis carried out by humans. Fourth, we discuss some additional rules, which, however, have less precise predictions and complicate the rule set. Fifth, we propose the Word Onset Hierarchy, which demonstrates the role of mobile vowels in breaking up sequences of identical segments and helps us clarify the relative importance of place of articulation, manner of articulation and voicing for mobile vowels. Finally, we argue that mobile vowels are gradient in nature and represent an asymmetric relationship between *v* and *vo*.

[2] Mobile vowels (Russian: *beglye glasnye*) are also known as "fleeting vowels" (e.g., Klapper 1993). In theoretical linguistics, they are often referred to as "yers" or "jers" (e.g., Scheer 2011 and 2019).

Our argument is structured as follows. In section 2, we present the machine learning experiment and show how the results can be translated into a set of simple rules. Section 3 evaluates the rules from the perspective of language pedagogy and identifies three issues that are explored in sections 4 through 6. In sections 7 through 9 we discuss the general properties of mobile vowels, before we summarize our findings and discuss their implications for language pedagogy in section 10.

2. A machine learning experiment

In order to investigate the motivation for the choice between *v* and *vo*, we created a database of 188,831 tokens extracted from the Russian National Corpus (main corpus).[3] We annotated the wordform following the preposition with regard to the following variables:

(3) Variables considered:
  a. First, second and third letter of the word
  b. Number of consonants before the first vowel (0–4)
  c. Type of vowel
  d. Place of articulation of the consonants before the first vowel
  e. Number of syllables
  f. The first vowel of the word

From the dataset we extracted 18,207 unique preposition-word pairs, on which the analysis was conducted. The variables in the dataset were encoded into one-hot vectors that were used in the models.

An overparametrized neural network was trained on a randomly drawn sample consisting of 90% of the dataset and then tested on the remaining 10%. Although the neural network is a “black box” and serves no purpose in terms of deriving grammatical rules, it enabled us to assess the explanatory potential of the chosen variables when it comes to empirically modelling the grammatical relationship under scrutiny in the present article. The reasoning behind this is that if there is a relationship to be found between the dependent and independent variables, a neural network would be able to infer it no matter how complex it is. The neural network consisted of three dense layers with 256 neurons each and RELU activation functions, as well as a final output layer with 1 neuron and a sigmoid activation function. Binary cross-entropy was used as the loss function and Adam as the optimizer. The network was trained with a batch size of 32 for 10 epochs. The hyperparameters of the network were not finetuned to increase accuracy but were simply chosen arbitrarily to create a network that would have sufficient explanatory capacity to model any potential grammatical relationship in the data.

[3] Our database is avaialble via TROLLing (Nesset, Xavier 2023).

When evaluated on the test set, the neural network model gave correct predictions for 97.4% of the preposition-word pairs. This shows that the network is able to learn to choose between the two variants of the preposition based on variables that reflect the environment in which they occur. Although the neural network will not play an important role in the further development of our analysis, the network represented a valuable first step, since the analysis gave us sufficient evidence that it was in fact possible to develop a rule set for mobile vowels based on the selected variables.

A binary decision tree classifier with a maximum depth of 2 was then estimated on a randomly drawn sample of 90% of the data. The model type and depth were chosen to increase simplicity of the resulting decision tree such that it could be used to create an understandable set of rules that a language learner or teacher could make use of. Binary decision trees are the simplest and most straightforward classifier of this type, and thus allow for both easier rulesets to be created and for them to be more easily interpreted. The maximum depth of the tree was set to 2 since the number of nodes in the tree increases exponentially with depth, which thus increases the complexity of the ruleset and makes it considerably less applicable in pedagogical situations. Gini impurity (a measure of misclassification) was used as the support criterion for the splits, and although – as shown in Table 1 – the dataset was significantly skewed towards *v* (93 % *v* vs. 7 % *vo* for the training data, 94 % vs. 6 % in the test data), classes were weighted equally. The model we created had an accuracy of 96.8% on the training data, and 96.7% on the test data.

	# *v*	**# *vo***	**# *v* + *vo***	**% *vo***
Training data	15,272	1,115	16,387	7 %
Test data	1,712	108	1,820	6 %

Table 1: Distribution of* v *and* vo *in training data and test data.

The results can be visualized as the annotated decision tree in Figure 1. Each branch in the tree includes four lines of text. The top line represents the relevant variable, which the model describes as X followed by a number in parentheses. This is followed by “<= 0.5” which represents the threshold where the variable is split. (Since all variables are indicator variables this will always default to 0.5.) For the convenience of the non-initiated reader the relevant variables are identified in the orange callouts.

The second line in each node displays the Gini impurity of the samples at the node, which represents the probability that a sample in the node would be mislabeled, were it labeled randomly according to the classification distribution at the node.

The third line, which is marked as “samples”, describe the total number of samples assigned to the relevant node. In the top node, the number indicates that the entire training set consisted of 16,387 samples.

The bottom line labeled "value" reports the distribution of the classified samples at each node. The number to the left is the number of examples with *v*, while the number to the right is the number of examples with *vo*.

The decision tree can be read as follows. As shown in the top node, the model first asks if the first letter of the wordform following the preposition is *v*. If the answer is "yes", we follow the arrow to the right and ask if the wordform starts with a single consonant or not. If the answer is "yes", the model predicts that the preposition has the form *v*. As shown, in the rightmost terminal node, this prediction holds for 579 out of 586 examples of words starting with the letter *v* followed by a vowel, i.e., words like *vokzal* 'railway station' and *vosem'* 'eight'.

If the word after the preposition starts with the letter *v* followed by a consonant, the model predicts that the preposition will have the form *vo*. This prediction is borne out by 608 out of 626 examples, as shown in the second terminal node from the right. Relevant examples include *vtoroj* 'second' and *vstreča* 'meeting'.

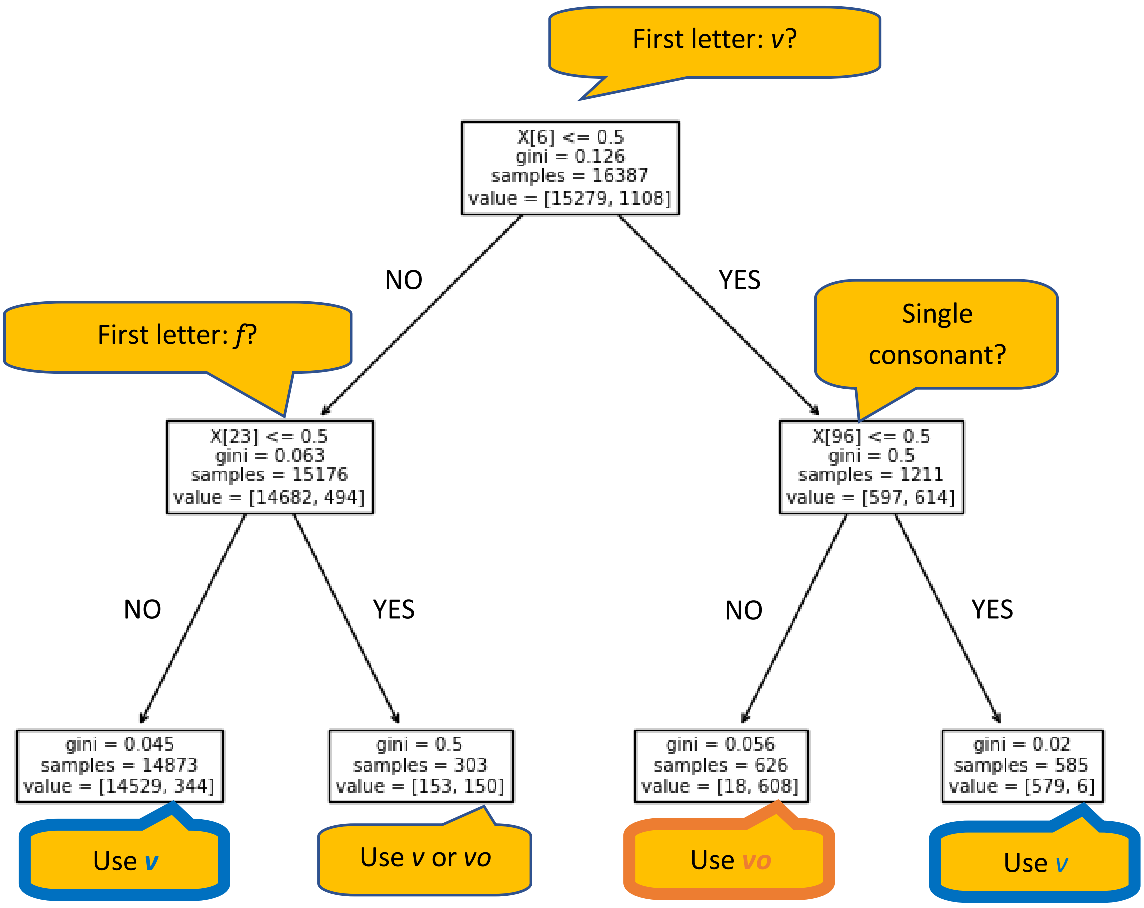

***Figure 1: Decision tree for the choice between* v *and* vo. The phrase "use *v* or *vo*" indicates that the model is not able to predict the choice of form, although only one option may be correct in each case.**

We now turn to the left portion of the decision tree, which concerns words that do not begin with the letter *v*. Here, the model asks if the word after the preposition begins with the letter *f*. If the answer is "yes", the model is not able to yield a clear prediction, since for such words there are 153 examples where the preposition is *v* and

150 where it is *vo*. This is shown in the second terminal node from the left, which covers words like *fabrika* ‘factory’ and *Francija* ‘France’.

The leftmost terminal node concerns all remaining examples, i.e., all examples where the word after the preposition starts with any other letter than *v* or *f*. This is the majority of the data. The model predicts that the form *vo* of the preposition, which is correct in 14,529 out of 14,873 examples.

The analysis visualized in the decision tree may be translated into a simple set of rules:

(4) Rules based on machine learning experiment:
  a. Before words beginning with *v* + a consonant, use *vo*.
  b. Before words beginning with *f*, use *v* or *vo*.[4]
  c. Elsewhere, use *v*.

Rule (4c) shows that *v* is a “normal case default” (Fraser and Corbett 1997), since its use is regulated by a general rule that applies unless there is a specific reason to apply another, more specific rule.

To scholars familiar with the literature on mobile vowels in Russian, the generalizations in (4) come as no surprise. In particular, it is well known that *vo* is frequently attested before consonant clusters beginning with *v*, a fact that is mentioned in grammars and textbooks (see, e.g., Timberlake 2004: 177–179). This shows that the machine learning approach is on the right track since it is able to discover well-known generalizations.

To what extent can the rules in (4) be helpful in the classroom? The set is simple in the sense that there are only three rules, and each rule is simple in the sense that they do not involve technical vocabulary that may be unknown to the average undergraduate student of Russian as a foreign language. The rule set furthermore allows the student to choose the correct form of the preposition in the vast majority of cases, at least if we assume that our dataset is representative of the Russian language as a whole. We submit that this is not a very controversial assumption, since we analyze a large dataset from a curated corpus. We therefore conclude that machine learning based on neural networks may represent a valuable tool from the perspective of language pedagogy. However, because of limitations with the machine learning algorithm used, the set of rules described above is not optimal with regard to either maximum accuracy, complexity, or applicability. In sections 3 through 6, we will identify three issues with the proposed rule set. We will also suggest further improvements to it in order to reduce the number of exceptions.

[4] In the same way as in Figure 1, the phrase “use *v* or *vo*” indicates that the rule cannot predict the choice between the two forms, although only one form may be correct in each individual case.

3. Evaluating the rules from the perspective of language pedagogy

In order to evaluate the rules in (4), it is instructive to consider the data in Table 2, which concerns words beginning with two consonants. The leftmost column indicates the first consonant of the cluster. The columns marked with a hashtag (#) offer raw numbers for *v* and *vo*, and the rightmost column gives the proportion of *vo* in percent. As shown in the table, some consonants are rare in consonant clusters, but for most consonants we have enough data to carry out a meaningful analysis.

First consonant	**# *v***	**# *vo***	**# *v* + *vo***	**% *vo***
b	193	19	212	9 %
v	19	665	684	97 %
g	247	16	263	6 %
d	213	18	231	8 %
ž	4	0	4	0 %
z	111	15	126	12 %
k	340	4	344	1 %
l	8	14	22	64 %
m	48	55	103	53 %
n	16	1	17	6 %
p	834	14	848	2 %
r	12	14	26	54 %
s	904	27	931	3 %
t	239	1	240	0 %
f	8	162	170	95 %
ch	56	7	63	11 %
c	26	1	27	4 %
č	20	3	23	13 %
š	85	1	86	1 %

Table 2: Distribution of* v *and* vo *before words beginning with two consonants. The leftmost column indicates the first consonant of the cluster.

Table 2 shows that for most consonant clusters *v* is the dominant option, since for most consonants the proportion of *vo* is less than 15%. This is what we would expect from the rules in (4), where *v* is the normal case default. At the other end of the scale, clusters beginning with *v* has 97% *vo*. This is also expected, insofar as rule (4a) states that *vo* is selected before a word that begins with *v* followed by a consonant.

However, Table 2 also contains data that are not expected from the rules in (4). First, consonant clusters beginning with *f* yields 95% *vo*, although rule (4b) and the data in Figure 1 indicate no preference for *vo* over *v* before words beginning with *f*. Clearly, therefore, the distribution of *v* and *vo* before *f* merits a closer look. We return to this “*f* problem” in section 4.

The rules in (4) do not mention clusters beginning with the labial nasal consonant *m*, nor are clusters in the oral sonorants *l* and *r* mentioned. This means that the default rule in (4c) will apply. The rules therefor predict *v* in front of these consonants. However, Table 2 shows that clusters with *m, l* and *r* have more than 50% *vo*. It is therefore necessary to consider the "*m* problem" and the "*l/r* problem" in sections 5 and 6.

4. The *f* problem

As shown in sections 2 and 3, the machine learning experiment did not yield clear predictions for words beginning in *f*. However, if we distinguish between words beginning with *f* + a vowel such as *fabrika* 'factory' and words beginning with *f* + a consonant such as *Francija* 'France', a much clearer picture emerges. As shown in Table 3, *v* is strongly preferred before *f* + a vowel – here we have only 2% *vo*. Before *f* + a consonant, on the other hand, *vo* dominates with 95% of the relevant examples in our dataset. This is expected from existing scholarly literature, including grammars and textbooks. For instance, Timberlake (2004: 179) points out that the mobile vowel "dissimilatively separates consonants that are similar in place and/or manner of articulation", such as *v* and *f*.

	#v	**#vo**	**# *v* + *vo***	**%vo**
f + vowel	159	3	162	2
f + consonant	8	162	170	95

***Table 3: Distribution of* v *and* vo *before words beginning with* f**

In section 2 we saw that *v* is used before words beginning in *v* + a vowel (e.g., *vokzal* 'railway station'), while *vo* is the preferred option when the preposition occurs before a word in *v* followed by a consonant (e.g., *vtoroj* 'second'). Table 3 indicates that words beginning in *f* behave the same way. In order to incorporate this insight in our analysis, we may revise the rules in (4). Instead of having separate rules for words in *v* and *f*, we may capture the uniform behavior of *v* and *f* in one rule. This makes it possible to dispose of rule (4b). In other words, we need only two rules to account for the distribution of *v* and *vo*.

(5) Revised rules after solving the *f* problem:
  a. Before words beginning with *v* or *f* + a consonant, use *vo*.
  b. Elsewhere, use *v*.

The rules in (5) have two advantages over those in (4). First of all, the rules in (5) improve the precision of the analysis, since they yield correct predictions for nearly all words in *f*. Second, the rule set in (5) is simpler than the one in (4), insofar that it refers to the same number of segments (*v* and *f*) with fewer rules. Further simplification of rule (5a) is possible if we make use of distinctive features. Since /v/ and /f/ represent a natural class of segments, we may state that *vo* is used before a labial fricative followed by a consonant.[5] A rule with distinctive features may work well for students with basic training in phonetics and phonology, while the version in (5a) may be preferrable for students without such training. Either way, our discussion of the *f* problem shows that the rules that emerged from the machine learning experiment are not optimal, since it is possible to advance an alternative that is both more precise and simpler. However, the rules from the experiment represent a good starting point for further analysis.

5. The *m* problem

The *m* problem is that the rules that emerged from the machine learning experiment predict *v* before all words in *m,* while the data reviewed in section 3 show that *vo* is attested in 53% of the examples with *m* followed by a consonant. Is it possible to improve the rule system so as to provide a better analysis of words in *m*?

Before words where *m* is followed by a vowel (e.g., *Moskva* ‘Moscow’), *v* is always used, so in the following we will focus on words with *m* as the first member of a consonant cluster. Examples include *mgla* ‘darkness’, *mnogo* ‘many’, and *mladšij* ‘youngest’. Such words are known to be tricky when it comes to mobile vowels in a preceding preposition (see, e.g., Timberlake 2004: 179 for discussion).

Table 4 summarizes the distribution of *v* and *vo* before nine different consonant clusters that are attested in our dataset. For some clusters, we have few examples, so it is necessary to group the clusters into broader categories. We start with clusters involving a fricative or an affricate, i.e., *mc* (*Mcensk* ‘name of town’*, mš* (*mšara* ‘mossy swamp’)*, mšč* (*mščenie* ‘revenge’), *mx* (*mxat* ‘name of theater in Moscow’), and *mz* (*mzdoimstvo* ‘bribery’). If we summarize the results for these consonant clusters, we observe a strong preference for *vo* (12 attestations) over *v* (1 attestation) in our dataset.

For clusters where *m* is followed by a plosive, the only example we have is the cluster *mg*, which has 3 attestations of *v* and 5 of *vo*. Here we have very little data, and no clear pattern emerges from the few attestations we have.

Clusters involving the two oral sonorants *l* and *r* (e.g., *mladšij* ‘youngest’ and *mračnyj* ‘dark’) show a fairly strong preference for *v*. For *ml*, the *v* to *vo* ratio is 11 to 4, while for *mr* the ratio is 15 to 1.

For the cluster *mn*, the situation is less clear. Table 4 reveals a preference for *vo*, but *v* is also widely attested. The majority of examples involve the root *mnog*- ‘many’ (37 out of 50 attestations), but even for these words no clear pattern emerges. There seems to be considerable variation, and both variants of the preposition appear to be

[5] We will not discuss the question as to whether /v/ should be analyzed as a fricative in Russian, since this is beyond the scope of the present study. In the same way as fricatives, /v/ undergoes devoicing. Thus, *v ten’* ‘into the shadow’ begins with a voiceless sound. Unlike fricatives, however, /v/ does not trigger voicing, as demonstrated by minimal pairs like *dvoix* ‘two (genitive)’ and *tvoix* ‘your (genitive plural)’.

acceptable to at least some language users. The amount of variation before the *mn* cluster is not unexpected. For instance, Timberlake (2004: 178–179) reports on considerable variation in this environment.

Cluster	**# *v***	**# *vo***	**# *v* + *vo***
mc	0	4	4
mš	0	2	2
mšč	1	0	1
mx	0	5	5
mz	0	1	1
mg	3	5	8
ml	11	4	15
mr	15	1	16
mn	18	32	50

Table 4: Distribution of* v *and* vo *before* m + *consonant

How can we summarize the situation? Is it possible to refine the rules in (5) and make our analysis more precise? An option would be to add a rule along the following lines:

(6) The *m* rule:
Before words beginning with *m* + a consonant, use *vo*.
Exception: Before words beginning with *m* + *l* or *r*, use *v*.

This rule is an improvement in the sense that it yields correct predictions for the majority of clusters in *m*. This is shown in Table 5, which provides an overview of the accuracy and number of exceptions for the various versions of the rule set under scrutiny. As shown in the rightmost column, the m-rule (including the "exception" concerning *m*+*l* and *m*+*r*) brings the accuracy up to 98.7% and the number of exceptions down 242.

	Decision tree	*f*-rule	*m*-rule	*m*-rule with exception (*m*+*l*, *m*+*r*)
Accuracy	96.8%	97.7%	98.5%	98.7%
Number of exceptions	577	423	273	242

Table 5: Accuracy and number of exception words for each proposed rule set

But this improvement comes at the cost of increased complexity. Not only do we have to add a rule to our rule set – we are also forced to accept the additional complexity of adding an exception to the rule. The question is whether it is worth it from the perspective of language pedagogy. It is instructive to compare with the *f* problem discussed in the previous section. For clusters in *f*, we were able to increase the precision of the analysis and at the same time simplify our rule system. Such a modification is clearly valuable. For clusters in *m*, the situation is less clear, and we are forced to weigh the benefits against the cost, a process that necessarily involves a degree of subjectivity. In addition to increased complexity, the *m* rule comes with two drawbacks. First, the rule covers relatively few examples since most clusters with *m* followed by a context are not frequent. A second disadvantage is the fact that there is considerable variation between *v* and *vo*, especially for words beginning in *mn*, as shown above. For such words, the *m* rule is not helpful, thus reducing its overall value.

In summary, there seems to be now clear answer as to whether we should add the *m* rule in (6) to our rule system. The decision may depend on the target group. While advanced students may find the *m* rule useful, for students at lower levels it might only be a source of frustration.

6. The *l/r* problem

We now turn to clusters beginning with the oral sonorants *l* and *r*, which are attested in words like *l’gota* ‘privilege’ and *rvanyj* ‘ragged’. The rules emerging from the machine learning experiment do not mention these clusters, and therefore predict the default variant of the preposition, i.e., *v*. However, as pointed out in section 2, *vo* is attested in more than 50% of the examples in our dataset where *l* or *r* is followed by a consonant.

We first consider clusters beginning with *l*. As shown in Table 6, we have few examples, and no clear pattern seems to emerge from the available data. This is not surprising, insofar as earlier students of these words have found considerable variation (see, e.g., the overview in Klapper 1993: 20–23).

Cluster	**#*v***	**#*vo***	**#*v* + *vo***
l + plosive	2	5	7
l + nasal	0	1	1
l + fricative	6	8	14

Table 6: Distribution of* v *and* vo *before* l + *consonant

In order to make explicit that both variants of the preposition appear before consonant clusters beginning with *l*, we may add the following rule to our set of rules:

(7) The *l* rule:
Before words beginning with *l* + a consonant, use *v* or *vo*.

The situation for clusters beginning with *r* is summarized in Table 7. We have only three clusters, *rv* (rvanyj ‘ragged’), *rž* (*ržanoй* ‘rye’), and *rt* (*rtu* ‘mouth (locative singular)’), and for each cluster we have little data. It is therefore not possible to draw strong conclusions. The table furthermore does not reveal a clear tendency. It seems that both *v* and *vo* are used before clusters beginning with *r*. It is possible to incorporate this insight by adding the following rule:

(8) The *r* rule:
Before words beginning with *r* + a consonant, use *v* or *vo*.

However, from the perspective of language pedagogy, adding this rule is at best a marginal improvement, given that it covers few examples and does not give any clear predictions. For this reason, it is even problematic to use the term “rule” about the statements in (7) and (8).

Cluster	***#v***	***#vo***	***#v + vo***
rv	6	4	10
rž	5	8	13
rt	0	2	2

Table 7: Distribution of* v *and* vo *before* r *+ consonant

7. Exceptions and the labial generalization

We now turn to what we may call “exceptions”, i.e., cases where the rules emerging from the machine learning experiment yield incorrect predictions. In particular, we are interested in examples where the neural network predicts *v*, although in actual reality *vo* is attested. Notice that we go back to the original rules in (4). As we shall see, this enables us to formulate a generalization about the likelihood of *vo* and the phonological properties of the following word.

Table 8 summarizes the situation. The table covers word onsets (word-initial consonant clusters or vowels) where the neural network yields incorrect predictions. Since we are interested in robust tendencies, the table only contains word onsets with more than 10 attestations in our dataset.

Word onset	***#attestations***
fr	113
fl	47
mn	32
i	28
o	20
bl	17
dv	17
u	12
gr	11

Table 8: Number of attestations where the neural network yields incorrect predictions. The table contains word onsets with more than 10 attestations in our dataset.

We have already discussed the *fr* and *fl* clusters in section 4 and the *mn* cluster in section 5. These three clusters share one property, namely that they begin with a labial consonant. The next consonant cluster in the table, *bl*, also begins with a labial consonant, while *dv* further down has a labial consonant as its *second* member. The generalization thus emerges that clusters containing a labial consonant are more likely to combine with *vo* than other consonant clusters. The only cluster in Table 8 that does *not* involve a labial consonant is *gr*, which has only 11 attestations in our dataset.

Turning now to the three word onsets involving vowels in Table 8, we see that two of them, *o* (e.g. *oružie* ‘weapon’) and *u* (*uslaždenie* ‘delight’), are rounded vowels. Since rounded vowels involve a labial articulation, these words lend additional support to the generalization about *vo* before labial word onsets.

8. The Word Onset Hierarchy

Our findings so far can be summarized as the following hierarchy which ranks word onsets according to their likelihood to combine with *vo*:

(9) The Word Onset Hierarchy:
*v*C, *f*C > *m*C > *l*C, *r*C > other labial
(where C stands for any consonant and > means “is more likely to be preceded by *vo* than”)

As we have seen, *vo* is most likely to occur in front of consonant clusters starting with *v* or *f*. Then follows clusters beginning with *m*, which in turn are followed by clusters in *l* and *r*. Mobile vowels are attested least consistently before other word onsets involving labial sounds.

In the theoretical literature, mobile vowels have often been analyzed as a so-called OCP (Obligatory Contour Principle) effect (e.g., Linzen, Kasayanenko and Gouskova 2013). OCP is a constraint that bans sequences of identical or nearly identical segments. The Word Onset Hierarchy lends support to an analysis along these lines, since clusters beginning with *v* (and its voiceless counterpart *f*) are most likely to combine with *vo*. In other words, the addition of the mobile vowel prevents us from having *v* immediately followed by *v*.

At the same time, the Word Onset Hierarchy enables us to clarify some additional properties of mobile vowels. First, the mobile vowel in *vo* primarily occurs in front of consonant clusters. *Vo* is marginal before words beginning with a vowel, and it is generally not found before words beginning with a single consonant.[6]

A second point concerns the contrast between voiced and voiceless segments. The Word Onset Hierarchy suggests that voice is not a relevant feature, since clusters with the voiceless *f* behaves the same way as clusters with the corresponding voiced *v*.

Third, the Word Onset Hierarchy indicates that place of articulation is a relevant factor. As we have seen in the previous section, *vo* is first and foremost attested before labial segments. However, as shown in section 6, *vo* also occurs before clusters beginning with *l* and *r*, which have dental (or alveolar) place of articulation. We

[6] The combination of *v* followed by a word in a single *v*, e.g., *v Vene* ‘in Vienna’, is normally pronounced with a geminate (long) consonant.

speculate that the somewhat inconsistent use of *vo* before *l* and *r* may be due to the non-labial place of articulation of these sounds.

Fourth, manner of articulation also seems to play a role. The Word Onset Hierarchy is dominated by sounds without oral closure, such as fricatives, nasals and oral sonorants. Plosives (oral stops), on the other hand, typically do not trigger the mobile vowel in *vo*. Even the voiced labial plosive *b* is only marginally attested in combination with *vo*, as shown in section 6. This finding may be related to sonority. Plosives typically occur in clusters with rising sonority such as *bl* and *pr*. It has been argued in the scholarly literature that mobile vowels are most likely to appear before clusters with falling sonority (Linzen, Kasayanenko and Gouskova 2013: 455).

A final point emerging from the Word Onset Hierarchy concerns variation. Some scholars have argued that rules for mobile vowels may not be categorical. For instance, Linzen, Kasayanenko and Gouskova (2013: 455–457) uncovered "stochastic phonological constraints and found that the lexical variation is much more extensive than previously known". Our findings lend support to this, since the Word Onset Hierarchy indicates that mobile vowels, at least for the preposition under scrutiny in the present study, are not an all or nothing affair. Even for the consonant clusters that are most likely to trigger the mobile vowel in *vo*, we find occasional examples with *v*. As we move from left to right in the hierarchy, the amount of variation increases. We will elaborate on the relevance of variation for mobile vowels in the following section.

9. Variation: an asymmetric relation

So far, we have been concerned with word forms that show consistent behavior in our dataset. Either they combine with *v*, or they are preceded by *vo*. We may call such words "rational". In what follows, we will consider "irrational" words, i.e., words that show inconsistent behavior, insofar as they combine with both *v* and *vo* in our dataset. As we will see, irrational words shows that the relationship between *v* and *vo* is asymmetric.

We can divide "irrational words" into three groups: those where *v* is more frequent than *vo*, those where *v* and *vo* have the same frequency, and those where *v* is less frequent than *vo*. Table 9 summarizes the situation. As shown, cases of the first type, where *v* is the most frequent variant, are very rare. In our dataset, only one wordform (type) is attested, the genitive/locative plural form *mnogix* of *mnogo* 'many'. This form has a total of 4 attestations (tokens), 3 with *v* and 1 with *vo*.

The second group, where the two variants of the preposition are equally frequent, is also relatively rare in our dataset. We have 153 wordforms (types), all of which are represented with 1 attestation of *v* and 1 of *vo*.

The third group dominates our dataset with 232 wordforms (types) and more than 23,000 tokens. For all these types, there is only 1 token with *v*, while *vo* is represented with up to 5,745 tokens (for *vtoroj* ‘second’).

	# Types	# Tokens
v > vo	1	4
v = vo	153	306
v < vo	232	23,234
Total	386	23,544

Table 9: The distribution of “irrational words” that combine with both* v *and* vo *in our dataset.

The data in Table 9 lend further support to the idea from the previous section that mobile vowels are not an all or nothing affair. There is considerable variation, and even words like *vtoroj*, which has a consonant cluster beginning with a labial fricative and thus is expected to strongly prefer *vo,* may occasionally combine with *v* in corpus data. However, this variation is not random, since in most cases it concerns words that normally take *vo*. In other words, *vo* may occasionally be replaced by *v* in environments where *vo* is expected, while the opposite is almost never the case. The relationship between *v* and *vo* is therefore asymmetric. This lends support to our analysis from section 2, whereby *v* is the normal case default variant of the preposition.[7]

10. Conclusions and implications

This article reports on a machine learning experiment concerning mobile vowels in the preposition *v* ‘in(to)’. We have demonstrated that a neural network is able to correctly predict the occurrence or non-occurrence of mobile vowels in 97.4% of the cases in our dataset. A decision tree model was used to develop a set of three rules that nearly match the accuracy of the neural net.

We have argued that these rules may be valuable for students of Russian since the rules are simple and yield correct predictions in the vast majority of cases. However, we have argued that the rule set may be improved, and we have proposed a modified version that is both simpler and more precise. This suggests that machine learning may be a valuable tool in language pedagogy, but that the results from machine learning experiments cannot be taken at face value. Stated differently, machine learning supplements, but does not replace linguistic analysis carried out by human beings.

[7] As pointed out by an anonymous reviewer, we cannot exclude the possibility that the occasional attestations of *v* instead of *vo* are due to typos, since typos may occur even in thoroughly curated corpora like the Russian National Corpus. However, even if some of the relevant examples are typos, it is interesting that these typos (almost) always go in one direction (the use of *v* instead of *vo,* not the other way around). This testifies to the asymmetric nature of the relation between the two forms of the preposition.

Mobile vowels represent a substantial challenge for students of Russian as a foreign language. Our analysis has yielded a simple set of two basic rules, whereby (a) *vo* is used before consonant clusters beginning with *v* or *f*, while (b) *v* is used elsewhere. This rule set is so simple that it may be valuable for first year students. We have explored some additional rules for other consonant clusters. They have less precise predictions and complicate the set of rules, so their value for language pedagogy is less obvious. Arguably, these rules may be relevant for more advanced students.

Our analysis of the preposition *v* lends support to earlier studies where mobile vowels have been explained as OCP effects that prohibit sequences of identical or near-identical consonants. We have proposed a Word Onset Hierarchy, which clarifies several properties of mobile vowels. We have seen that mobile vowels appear before consonant clusters, and that the labial place of articulation of the first member of the cluster is more important than the manner of articulation (no oral closure), while voicing seems unimportant. Our analysis furthermore suggests that mobile vowels are not an all or nothing affair, and that the relationship between *v* and *vo* is asymmetric, insofar as *v* is the normal case default.

The present study is based on a small-scale machine learning experiment with one preposition. However, the promising results we have reported suggest that similar studies of mobile vowels in other environments may be a fruitful way to go in future research. It may also be valuable to apply machine learning to other areas of the Russian grammar that are challenging for students of Russian as a foreign language. On a more general level our study indicates that machine learning has the potential to become a useful tool for language pedagogy.

References

Becker, Michael & Maria Gouskova. 2016. Source-Oriented Generalizations as Grammar Inference in Russian Vowel Deletion. *Linguistic Inquiry* 47.3: 391–425.

Bethin, Christina. 1998. *Slavic prosody: Language change and phonological theory*. Cambridge: Cambridge University Press.

Blumenfeld, Lev. 2012. Phonology. Vowel–zero alternations in Russian prepositions: prosodic constituency and productivity. In: Veronika Makarova (ed.) *Russian Language Studies in North America: New Perspectives from Theoretical and Applied Linguistics*, pp. 43–69. London and New York: Anthem Press.

Fraser, Norman M. & Greville G. Corbett. 1997. Defaults in Arapesh. *Lingua* 103: 25–57.

Gouskova, Maria. 2012. Unexceptional segments. *Natural Language and Linguistic Theory* 30: 79–133.

Gouskova, Maria & Michael Becker. 2013. Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory* 31: 735–765.

Gribanova, Vera. 2009. Phonological evidence for a distinction between Russian prepositions and prefixes. In: G. Zybatow, D. Lenertová, U. Junghanns, and P. Biskup (eds.) *Studies in Formal Slavic Phonology, Morphology, Syntax, Semantics and Information Structure: Proceedings of FDSL 7, Leipzig 2007*, pp. 383–96. Frankfurt: Peter Lang.

Iosad, Pavel V. 2020. Per aspera ad astra. Nuli i zvezdočki v russkoj morfonologii. In: A.A. Kibrik, Ks. P. Semënova, D.V. Sičinava, S.G. Tatevosov, A.Ju. Urmančieva (eds.) *VAProsy jazykoznanija. Megaspornik nanostatej*, pp. 69–73. Moscow: “Buki Vedi”.

Klapper, John. 1993. Fleeting prepositional o in the contemporary Russian language. *Russian Language Journal* 47: 17–33.

Linzen, Tal, Sofya Kasyanenko & Maria Gouskova. 2013. Lexical and phonological variation in Russian prepositions. *Phonology* 30: 453–515.

Matushansky, Ora. 2002. On Formal Identity of Russian Prefixes and Prepositions. *MIT Working papers in Linguistics* 42: 217–253.

Nesset, Tore; Xavier, Kevin, 2023, "Replication Data for: From machine learning to classroom learning: mobile vowels and the Russian preposition v ‘in(to)’", https://doi.org/10.18710/ZCDX1B, DataverseNO, V1

Scheer, Tobias. 2011. Slavic yers. In: Marc van Oostendorp, Colin Ewen, Elizabeth Hume, and Keren Rice (eds.) *The Blackwell companion to phonology*, pp. 2936–2962. New York: Wiley-Blackwell.

Scheer, Tobias. 2019. On the Difference between the Lexicon and Computation (Regarding Slavic Yers). *Linguistic Inquiry* 50.1: 197–218.

Steriopolo Olga. 2007. Yer Vowels in Russian Prepositions. In: Richard Compton, Magdalena Goledzinowska and Ulyana Savchenko (eds.) *Proceedings of Formal Approaches to Slavic Linguistics 15: The Toronto Meeting*, pp. 365–85. Ann Arbor: Michigan Slavic Publications.

Timberlake, Alan. 2004. *A reference grammar of Russian*. Cambridge: Cambridge University Press.