Interaction of Semantics and Morphology in Russian Word Vectors

Yulia Zinova, Anastasia Yablokova, Ruben van de Vijver
Heinrich-Heine-Universität
Universitätsstraße 1, 40225 Düsseldorf
zinova@hhu.de, anyab100@uni-duesseldorf.de, ruben.vijver@hhu.de

Abstract

In this paper we explore how morphological information can be extracted from fastText embeddings for Russian nouns. We investigate the negative effects of syncretism and propose ways of modifying the vectors that can help to find better representations for morphological functions and thus for out of vocabulary words. In particular, we look at the effect of analysing shift vectors instead of original vectors, discuss various possibilities of finding base forms to create shift vectors, and show that using only the high frequency data is beneficial when looking for structure with respect to the morphosyntactic functions in the embeddings.

Keywords: Russian, FastText, Embeddings, Morphology, Semantic Classes, Syncretism

1. Background

Learners of morphology, whether humans or machines, must be able to overcome the Zipfian distribution of words: A few words occur extremely frequently, and most words are very infrequent (Kodner, 2022; Guzmán, 2020). As a result, there are many words that a learner has to produce, although these words have never been encountered in the input. A question that arises is what allows a learner to do this (Guzmán, 2020; Ackerman and Malouf, 2013)?

Some proposals to achieve this have focused on the form-side of morphology (Ackerman and Malouf, 2013; Albright, 2010; Malouf, 2017). These proposals leverage implications among forms in a paradigm. For example, the Latin genitive form of King *regis* allows a learner to predict most other forms of the paradigm (the dative form *regi*, the accusative *regem* and the ablative *rege*), whereas the nominative form *rex* does not. So, if the learner knows the genitive form, they can use this knowledge to predict all other forms (see Albright, 2010, for an explanation of how this mechanism affected diachronic changes in Yiddish).

Yet, it is not clear whether language users really use forms to produce other forms (Nieder et al., 2021a,b,c). One reason is that in most languages it is not obvious what is the most informative form. Finnish, for example, has several forms that can be used to base other forms upon (Nikolaev et al., 2022b). Moreover, this focus on form alone neglects any role of semantics in predicting the meaning of words that have not been encountered.

Information about the semantics of words can be captured by embeddings, and can be used to investigate properties of paradigms that can be helpful to learners. Recent work on morphology used the information contained in embeddings to investigate specific properties of morphology. For example, Westbury and Hollis (2019) has used embeddings to investigate whether part-of-speech can be predicted from embeddings.

Embeddings are learned in an unsupervised manner from raw text and thus contain information about the distribution of words in a corpus (Bojanowski et al., 2017; Landauer and Dumais, 1997; Mikolov et al., 2013; Pennington et al., 2014). An additional difficulty for languages with rich morphology, such as Russian, is the huge amount of out-of-vocabulary words. This problem is addressed by fastText word vectors (Bojanowski et al., 2017), since the model learns representations of the n-grams and thus a representation of any given word is either directly learned or calculated on the basis of the n-gram representation of its parts.

FastText representations have been shown to work well for many tasks, however, there is potential for improvement. The idea of representing the word through the sum of its n-grams relies on the idea that affix n-grams correspond to functions and these functions can be represented in a similar way as the whole words. It has been shown by Nikolaev et al. (2022a) and Shafaei-Bajestan et al. (2022) that though this idealisation works, it is not accurate to represent a single morphosyntactic function with a unique vector: there is interaction between various functions (Nikolaev et al., 2022a) as well as between a function and semantic classes (Shafaei-Bajestan et al., 2022). An additional problem in Russian is syncretism: same affixes can be encountered in various cells within a paradigm as well as represent different functions across the paradigms. For example, the genitive and accusative of the word for an animate masculine noun 'elephant' are both *slona* and at the same time for example the nominative singular of an animate feminine noun 'mother' *mama* have the same affix. As a result, the embedding of the word *slona* necessarily contains distributional information about its occurrence in genitive and in accusative contexts and the representation of a trigram ending on *a* at the end

The Workshop on Cognitive Aspects of the Lexicon, pages 120–128
20 May, 2024.

of the word may, in addition, contain distributional information about nominative singular contexts.

In order to generate reliable predictions for the meaning of each morphosyntactic function of syncretic forms, we propose to extend a method proposed in (Nikolaev et al., 2022a; Shafaei-Bajestan et al., 2022). Instead of assuming that the meaning of a word is best predicted by the sum of the vectors of its n-gram components, we represent as a sum of the embeddings for the base and the morphosyntactic functions it expresses. In order to do so, we use dictionaries and existing morphological tools in order to find out-of-vocabulary word forms. We also use the word forms for which we have embeddings to extract vectors for lemmas and grammatical functions. We then use the inferred vectors to predict vectors for out-of-vocabulary word forms, to provide multiple vectors for syncretic forms. At the current stage we rely on dictionary information in order to explore such representation, but our final aim is a self-supervised pipeline.

According to Wiemerslage et al. (2022) the next challenge in computational morphology is to understand morphology from text alone. Wiemerslage et al. 2022 introduces the task of *truly unsupervised morphological paradigm completion* and proposes a pipeline for approaching it. In a first step, Wiemerslage et al. (2022) clusters word forms into paradigms on the basis of their orthographical similarity. In a second step, it is assessed which orthographic changes on the word forms express the same inflectional information. For example the last character in the Russian word *okna* 'windows' and the last one in *mamy* 'mothers' express the same inflectional information (namely, nominative plural). Information about word embeddings is then used to assess the distribution of such inflections, and this, in turn, is used to assign labels to word forms. These labeled word forms are then used to train a morphological learner. The model presented in Wiemerslage et al. 2022 is trained on digitized children's books and the Bible in several languages (German, Greek, Icelandic, and Russian). The evaluation has been done in terms of correct paradigm reconstructions with paradigm slots aligned between different lemmas but in random order, the best possible correspondence to true labels being selected for the evaluation. The best results across all the languages and training data are about 27% correctly generated word forms for Russian digitized children's books. The pipeline proposed in Wiemerslage et al. 2022 is in principle unable to cope with syncretism, since any string can be mapped to only one functional slot. This raises the question how morphology can be learned in an self-supervised way while also taking into account the fact that a lot of languages exhibit syncretism.

While the pipeline of the Wiemerslage et al. (2022) works with the original vectors, comparisons among vectors yields further vectors, and there have been proposals in the literature to look at the structure of such comparisons instead (Nikolaev et al. 2022a, Shafaei-Bajestan et al. 2022). For example, one could assume that the vector of one word form, which we refer to as the base vector, in a paradigm is used to derive other word forms. An obvious choice for the base vector would be the nominative singular form, since it is the base form provided in the dictionaries. But the nominative case often is syncretic in Russian, which becomes especially concerning when working with other forms, with which the nominative singular is syncretic. Furthermore, the same dictionaries often list a set of other forms (principal parts) to provide the full information needed to reconstruct the paradigm, which may serve as an indicator that nominative singular alone may not be enough for our purposes. In the following we will investigate various choices for a base vector.

2. Methodology

2.1. Data

To explore the semantic space of Russian nouns, we first need an overview of the nominal paradigms. We obtained it by extracting 14,157 nouns from a recent frequency dictionary (Ljaševskaja and Šarov, 2009) together with their frequency information. These nouns were parsed using the `pymorphy2` library (Korobov, 2015) and inflected along the list of fourteen forms: seven cases and two numbers. The cases include the six standard cases as well as the second genitive (partitive), here with the abbreviations used further throughout the paper: *nominative (nom), genitive (gen), dative (dat), accusative (acc), ablative (abl), locative (loc), genitive 2 (gen2)*. Each of these cases occurs in the *singular (sg)* or *plural (pl)*, as in Table 1.

We have excluded nouns for which *pymorphy* could not find a parse as well as those where not all of the paradigm sells were populated (this includes all pluralia tantum and all singularia tantum nouns as sell as nouns with paradigm gaps). After this, we were left with 11320 nouns, which amounts to 158480 forms. As one can see in Table 1, there are a lot of syncretic forms in the nominal paradigms, and this holds true for every paradigm type. In our dataset, these are 89738 forms, or 56,63% of the total number of forms. This leaves 68742 (43.37%) non-syncretic forms. This indicates that syncretism is a huge difficulty for learning of Russian morphology.

Case/Number	kniga	mama	čaj	slon	yabloko	mol’
	f, inan	f, anim	m, inan	m, anim	n, inan	f, anim
	book	**mother**	**table**	**elephant**	**apple**	**moth**
Singular						
Nominative	kniga	mama	čaj	slon	yabloko	mol’
Genitive	knigi	mamy	čaja	slona	yabloka	moli
Dative	knige	mame	čaju	slonu	yabloku	moli
Accusative	knigu	mamu	čaj	slona	yabloko	mol’
Instrumental	knigoj	mamoj	čajem	slonom	yablokom	mol’ju
Locative	knige	mame	čae	slone	yabloke	moli
Genitive 2	knigi	mamy	čaju	slona	yabloka	moli
Plural						
Nominative	knigi	mamy	čai	slony	yabloki	moli
Genitive	knig	mam	čaëv	slonov	yablok	molej
Dative	knigam	mamam	čajam	slonam	yablokam	moljam
Accusative	knigi	mam	čai	slonov	yabloki	molej
Instrumental	knigami	mamami	čajami	slonami	yablokami	moljami
Locative	knigax	mamax	čajax	slonax	yablokax	moljax
Genitive 2	knig	mam	čaëv	slonov	yablok	molej

Table 1: Nominal paradigms of Russian feminine inanimate (book), feminine animate (mother), masculine inanimate (tea), masculine animate (elephant), neuter inanimate (apple) and feminine animate of a different type (moth) nouns, annotated for case and number.

2.2. Word vectors

We created our own FastText vectors by training on a cleaned version of Russian Wikipedia using the cbow algorithm and otherwise standard settings. The obtained model provides vectors with 300 dimensions and is used in the visualizations and classification experiments presented in the following sections.

3. Visualising semantic space

In this section we present the main results of exploring the data through dimensionality reduction and visualization using principal component analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE, van der Maaten and Hinton 2008). In all of the following, the vectors of 300 dimensions were first reduced to 50 dimensions with PCA and then further reduced to two dimensions with t-SNE.

With the help of the visualizations, we explore three modifications that can help unveil the morphosyntactic functions of the vectors representing word forms: a comparison between the original and difference vectors, restricting the analysis to high frequency items, and varying the base for the computation of the difference vectors. We then see how these modifications affect the visualizations of both syncretic and non-syncretic forms.

3.1. Original vectors compared to difference vectors

The top part of Figure 1 visualizes the reduced original vectors of all the noun forms. For a better representation of syncretism and clarity, we label not individual forms, but all possible combinations of functions that can be expressed in one form. The legend of 1 is for both plots. Interestingly, although for fourteen forms 2^{14}-1 (16383) combinations are theoretically possible, only 40 are attested. We will call these combinations *case-number subsets*. The reduced semantic space has some discernible clusters, which means that some morphosyntactic functions occupy different areas in the semantic space. For example, the pink area on the top right (ablative plural words). At the same time, a lot of areas contain a mixture of vectors representing various case-number subsets.

Since it has been proposed in the literature to investigate the properties of vectors by studying shift vectors (these are vectors that represent the difference between two different forms, for example the difference between a plural and a singular; see Nikolaev et al. 2022a, Shafaei-Bajestan et al. 2022 for discussion), the bottom part of Figure 1 represents reduced difference vectors: for each noun form, a difference vector is obtained by subtracting a base vector from the vector for this form. In this figure the mean vector across all the forms of the paradigm is taken as a base vector.

The comparison of the original (top) and the difference (bottom) vectors reveals that the clusters based on the difference vectors are much clearer.

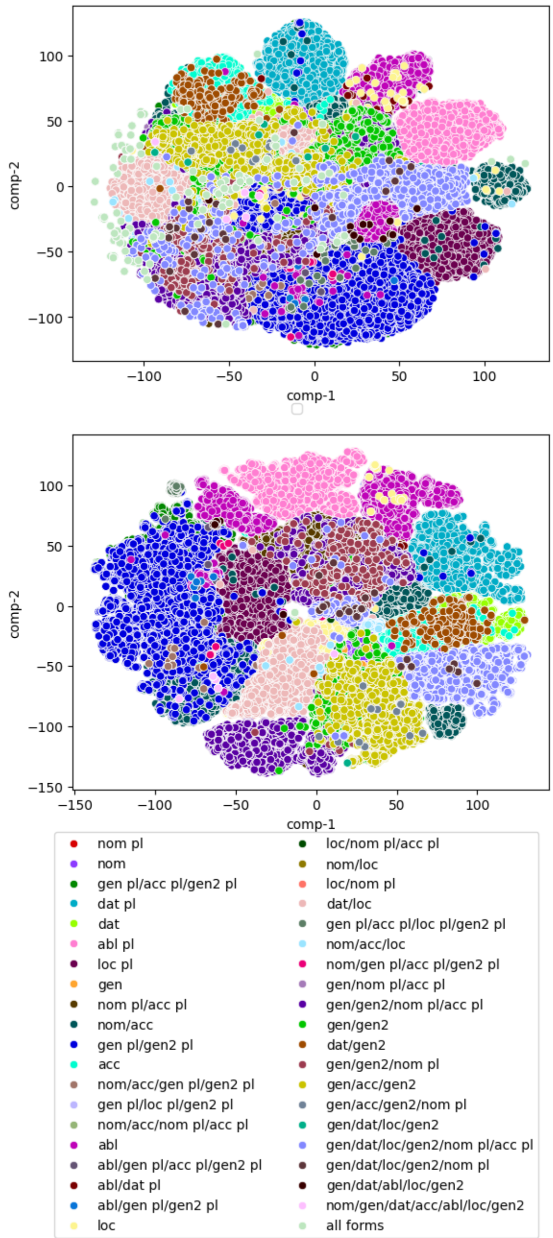

Figure 1: Visualization of nominal paradigms in Russian including syncretic ones. Original vectors are visualised on the top plot and difference vectors with the mean as a base vector on the bottom plot.

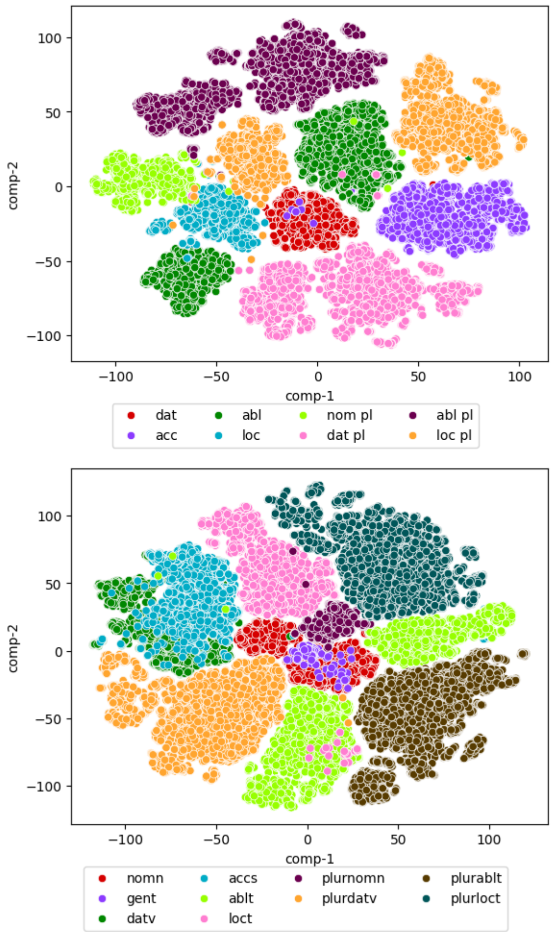

Figure 2: Visualization of nominal paradigms in Russian, only forms that are not syncretic, difference vectors. As a base, the nominative singular on top and the mean vector on the bottom.

Syncretism, however, clutters the semantic space, as is to be expected. The large number of case-number subsets makes visual analysis complicated and contributes to overlapping placement of vectors belonging to different subsets. In order to be able to further reduce the effect of syncretism, we have to investigate the properties of vectors of non-syncretic items.

3.2. Removing the syncretism

In order to explore the role of syncretism, we have removed syncretic forms from the analysis. The remaining 68742 forms are visualised as the original vectors (top) as well as difference vectors (bottom) in Figure 2. We observe clearer clusters for both the original and the difference vectors as well as the separation of the same case-number representations into multiple clusters.

3.3. The effect of frequency

A further idea that can be used to improve visualizations and later to approach the unsupervised learning is the option of limiting the input vectors to vectors of high frequency nouns. The reason for this is that forms of high frequency nouns occur more often in the data and thus are likely to have better representations. In addition, for high frequency lexemes we expect more forms to be encountered in the data and therefore have been learned by the algorithm and not constructed out of the n-gram representations.

The application of this approach is illustrated in Figure 3. The vectors for this visualization have

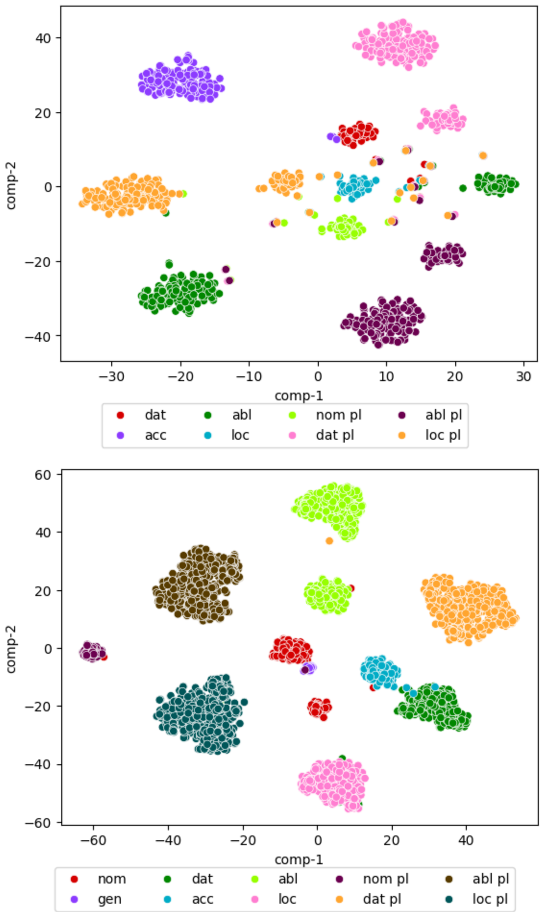

Figure 3: The syncretic forms were taken out by using dictionary information and the frequency restricted to 100 ipm or higher. Nominative singular as the base on the top and mean vector as the base on the bottom.

been filtered by frequency: only nouns with ipm of 100 and more according to Ljaševskaja and Šarov (2009). As can be seen, such restriction leads to more visually separated clusters of reduced non-syncretic vectors. Since frequency information can also be extracted during the learning process, such reduction of data can be useful in some steps of the pipeline for discovering morphological functions and their representations in a self-supervised manner.

4. Representing a lexeme

The given dictionary form is nominative singular, but is that form a good representation for the entire lexeme? For phonological reasons dictionaries also contain principal parts, which can be used to predict the phonological forms of other words in the paradigm (Albright, 2010; Nikolaev et al., 2022a). From the point of view of the embeddings, rather than the phonology of the forms, it is unclear whether principal parts are also needed to predict the semantics of other forms in the paradigm. We hypothesize that the mean vector over all forms of one lexeme represents the meaning of the lexeme better than a vector of any single form. It allows more uniform representations of morphological functions as shift vectors as well as the presence of a representation for all of the functions.

From the comparison of the top and the bottom plots in Figure 3 we can see that difference vectors are more useful when it comes to identifying clusters of vectors representing the same function, since on the bottom figure there are less morphosyntactic functions which are broken down into two distinct clusters. Let us now, after seeing the effect of frequency filtering, reintroduce syncretism, as is done in Figure 4. As above, we are considering the original vectors (top) and the difference vectors (bottom). Certain groups of forms are clearly identifiable on both visualizations: these are plural ablative or plural dative. The reason for this is that they are never syncretic (apart from the nouns that do not change their form at all, such as *mango* 'mango', absent on 4 due to the frequency restriction). As for the other number-case combinations, a couple of them form clusters without clearly identifiable borders, and most of them are spread out as in the top of figure 4. At the same time, although not all of the functional clusters can be clearly identified in the bottom of figure 4, the number of such clusters is significantly higher.

Interestingly, difference vectors representing a set of forms are often split into several clear clusters, which turn out to be related to different genders, as illustrated by 5. The most prominent example is the syncretic nominative/accusative singular form: in Figure 4 it is visible as three separate clusters. In 5 we see that these clusters correspond to the three genders, the neutral being positioned at the left periphery, the feminine between the feminine non-syncretic nominative and accusative and the masculine close to a (smaller and thus less visible) group of non-syncretic masculine nominative singular representations.

5. Supervised Classification

Although we work towards self-supervised learning, we ran a supervised classification task based on our vectors paired with morphological information. This had two goals: first, show that the data contained in the vector representations is enough in order to find all the 40 case-number sets, second to test the effect of proposed vector modifications. We built a Support Vector Machine to investigate to what extent the vectors are able to correctly use the meaning of different word forms to classify them. The words were split into 80% for training and 20% for testing. The accuracy of classification of all

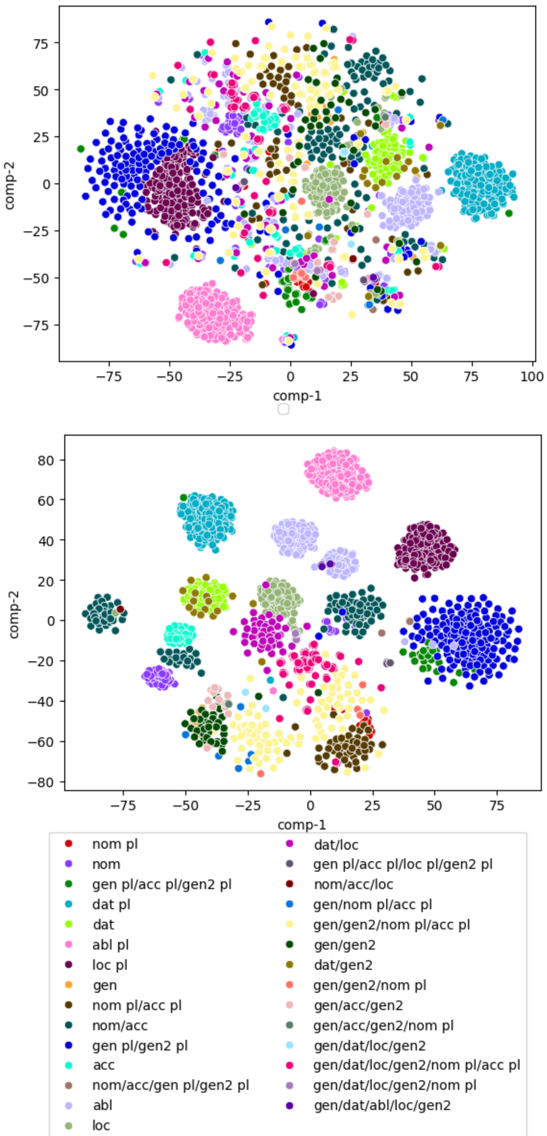

Figure 4: Original (top) and difference (bottom) vectors of relatively frequent Russian nouns including syncretic forms.

words (with syncretic forms) is 86.75%. As mentioned above, the classification needs to be done into one of the 40 categories corresponding to case-number set. This means that a word form is considered to be classified correctly only when the exact combination of morphosyntactic information is identified. So if a form is syncretic between nominative and accusative singular, it is correctly classified only if exactly this set of features is identified. This task is significantly harder than identifying one function from a set of functions a noun form can refer to.

We ran the same classification task for the subset of vectors representing non-syncretic forms. In this case the accuracy of the classification reaches 97.75%.

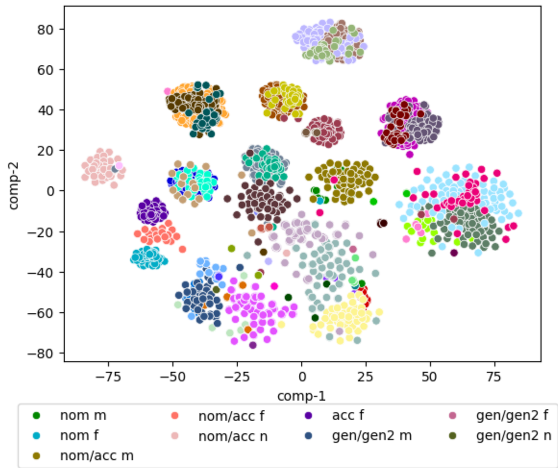

Figure 5: Difference vectors of relatively frequent Russian nouns including syncretic forms, colored by form and gender.

We have repeated both tasks for difference vectors instead of the original vectors. In both cases we see an improvement of classification: for all the forms the accuracy is 89.1% and for the non-syncretic forms it is 99.11% after this modification.

6. Semantic Classes

It is not surprising that the vector representations of nouns contain information about the semantic categories to which they belong. As we have seen, though, this information becomes less prominent if we perform a dimensionality reduction on the group of vectors that contain various case-number forms. To test the hypothesis that shift vectors correlate with the semantic class, we manually annotated 1576 nouns with 64 category labels, allowing each noun to receive multiple labels. Figure 6 represents nouns from 27 categories that are the most populous. It illustrates that if we run the PCA-tSNE reduction on the set of forms that are associated with one specific case-number function, we can observe the semantic grouping of the nouns, as illustrated by Figure 6. This result is in line with the findings of Shafaei-Bajestan et al. (2022). On the other hand, Figure 6 shows that there are no clear borders between the semantic categories and many of them get split into smaller clusters depending on their grammatical gender, as can be seen in Figure 7.

Separation between different classes becomes easier to follow if we restrict the number of classes included in the analysis, for example, to three classes that are expected to have distinct semantics, as shown in Figure 8.

Shafaei-Bajestan et al. 2022 have shown for English that the plural shift vector is not uniform across

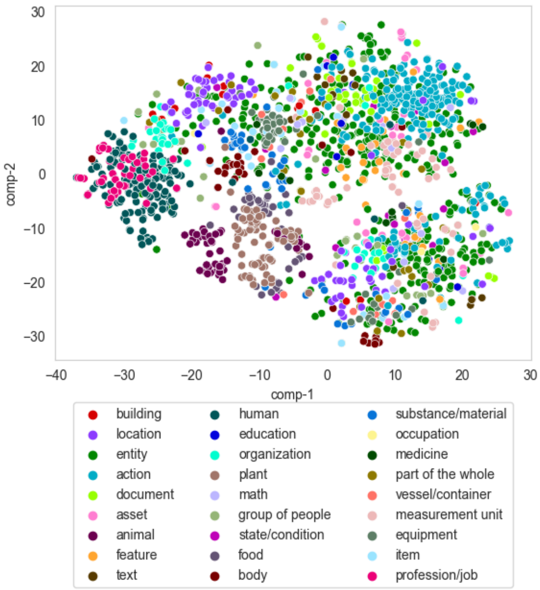

Figure 6: Original vectors of all semantic groups in dative singular

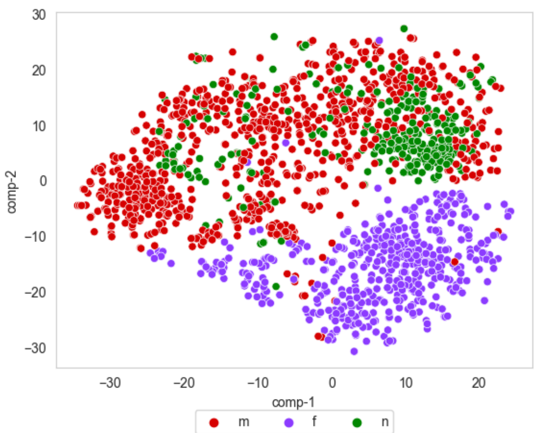

Figure 7: Original vectors of all semantic groups in dative singular colored by gender

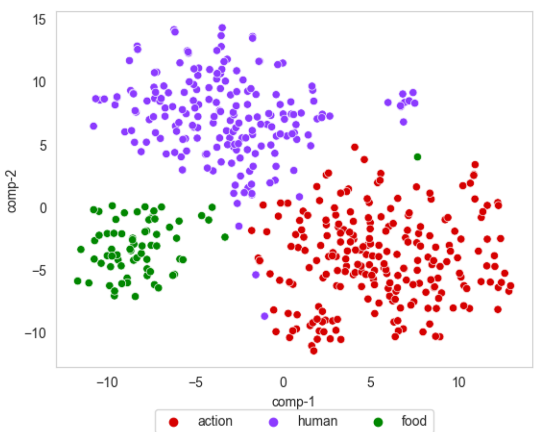

Figure 8: Original vectors of three big semantic groups in prepositional singular

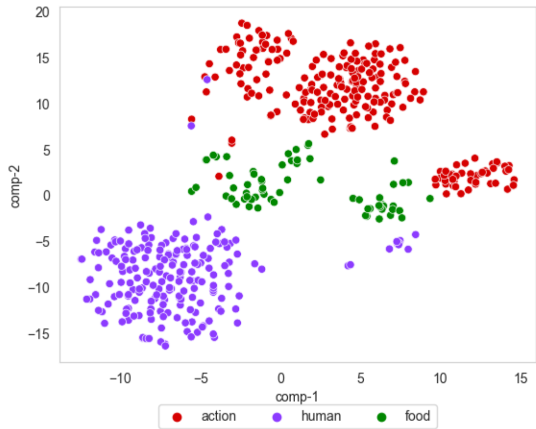

Figure 9: Shift vectors of three big semantic groups for dative singular and nominative singular as the base

semantic categories. One possible idealization would be to assume that an individual vector for a *noun* in some *form* would be a sum of the vector representation for the *noun* and the vector representation of the *form* (mutually independent). A candidate for the representation of the *noun* is the *nominative singular* representation of that noun if one assumes that nominative singular is an unmarked form and other forms are derived from it. This is the assumption in Shafaei-Bajestan et al. 2022, where the analyzed difference vectors are the difference vectors between the plural and the singular forms of a given noun. As in Russian there are many potential candidate forms when it comes to calculating a difference vector, we have explored all the possibilities of taking any given form as a base representation as well as taking the mean of all the forms of one lexeme as a base representation of the meaning of that lexeme. Our experiments have revealed that for any of the mentioned choices of the base form representation, the resulting difference vectors still carry semantic information about the class the noun belongs to. Several examples (with only three big semantic groups) are presented in 8 (original vectors) This allows us to conclude that morphosyntactic functions as learned by FastText correlate with the semantic class of the noun.

A comparison between the shift vectors with a nominative singular base, as in Figure 9, and the shift vectors with a mean base, as in Figure 10, show that the interaction between the semantic class and the morphosyntactic functions is present independently of the choice of the base form.

7. Updating representations

Based on the insights obtained from data visualisation and classifications, we have created base vector representations for nouns as well as rep-

	nom	gen	dat	acc	abl	loc	gen2	nom pl	dat pl	abl pl	loc pl	mean
all items	0.66	0.72	0.66	0.66	0.66	0.66	0.73	0.69	0.62	0.64	0.63	**0.78**
non-syncr	0.67	0.68	0.69	0.66	0.72	0.68	0.68	0.68	0.73	0.75	0.72	**0.80**
non-syncr in vocab	0.71	0.71	0.70	0.69	0.75	0.70	0.71	0.70	0.70	0.73	0.69	**0.82**

Table 2: Cosine similarity of constructed vectors with various base selection and fastText vectors

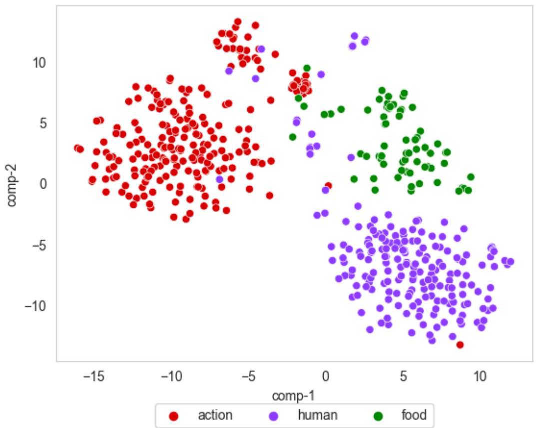

Figure 10: Shift vectors of three big semantic groups for dative singular and mean as the base

resentations for various functions and compared them with representations obtained thorough fastText. We then have compared how similar the resulting representations are to the original vectors in multiple conditions. First, we have tested every case as a potential base. For each base selection we have calculated the mean cosine similarity of all the items (first row of Table 2), the mean cosine similarity for non-syncretic items (second row) and the mean cosine similarity of non-syncretic items that are in vocabulary of the fastText model (last row). As is evident from the table, the mean as the base provided the best results, despite the fact that for each single case as a base all the items of that case would be identical to the original vectors, contributing similarity score of 1. Among the row comparisons we see that removing syncretic items and limiting the comparison to in vocabulary items increases the similarity. Based on the last observation we expect that replacing out of vocabulary representation with our constructed vectors will improve the performance of the model in downstream tasks.

8. Conclusion

We set out to investigate the effects of syncretism on learning Russian nominal paradigms from their embeddings. We are interested in doing this as a first step towards unsupervised learning of morphology. For this a pipeline has been proposed in Wiemerslage et al. (2021), but this pipeline did not take into account the effect of syncretism, which is very prevalent in Russian (our data set contained 43.4% non-syncretic forms).

We found several possible interventions that can be integrated into pipelines for semi-supervised or unsupervised learning of morphology. First, shift vectors provide a better basis for an analysis than original vectors, which is confirmed both by the visual analysis and the classification task results. The best choice of a base vector for obtaining the shift vectors, according to our observations, is an average vector of the paradigm. Since learning pipelines usually include a step of gathering forms of one paradigm, creating an average vector in an unsupervised manner should not cost additional problems and we hypothesize that in the absence of the labeled data this is the most robust choice.

We found that using high frequency items is beneficial for discovering structure in the data, both with and without syncretism. As is evident from the visual representations, in the latter case this modification is even more important and might help in the initial steps of the pipelines for unsupervised morphological learning. Although in this paper we still rely on labeled data for exploring the effect of the proposed vector modifications, we aim to leverage linguistic insights about morphological phenomena and use the resulting information to contribute to unsupervised learning of morphology.

9. Bibliographical References

Farrell Ackerman and Robert Malouf. 2013. Morphological organization: The low conditional entropy conjecture. *Language*, 89(3):429–464.

Adam Albright. 2010. Inflectional paradigms have bases too: Arguments from Yiddish. *Natural Language & Linguistic Theory*, 28(3):475–537.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Naranjo Matías Guzmán. 2020. Analogy, complexity and predictability in the russian nominal inflection system. *Morphology*, pages 1–44.

Jordan Kodner. 2022. Computational models of morphological learning. In *Oxford Research Encyclopedia of Linguistics*.

Mikhail Korobov. 2015. Morphological analyzer and generator for russian and ukrainian languages. In *Analysis of Images, Social Networks and Texts: 4th International Conference, AIST 2015, Yekaterinburg, Russia, April 9–11, 2015, Revised Selected Papers 4*, pages 320–332. Springer.

Thomas K Landauer and Susan T Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211.

Olga Nikolaevna Ljaševskaja and Sergej Aleksandrovič Šarov. 2009. *Častotnyj slovar sovremennogo russkogo jazyka na materialax Nacionalnogo korpusa russkogo jazyka*. Obščestvo s ograničennoj otvetstvennostju" Izdatelskij centr" Azbukovnik".

Robert Malouf. 2017. Abstractive morphological learning with a recurrent neural network. *Morphology*, 27(4):431–458.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 746–751, Atlanta, Georgia. Association for Computational Linguistics.

Jessica Nieder, Fabian Tomaschek, Enum Cohrs, and Ruben van de Vijver. 2021a. Modelling Maltese noun plural classes without morphemes. *Language, Cognition and Neuroscience*, 0(0):1–22.

Jessica Nieder, Ruben van de Vijver, and Holger Mitterer. 2021b. Knowledge of Maltese singular–plural mappings: Analogy explains it best. *Morphology*, pages 1–24.

Jessica Nieder, Ruben van de Vijver, and Holger Mitterer. 2021c. Priming Maltese plurals: Representation of sound and broken plurals in the mental lexicon. *The Mental Lexicon*, 16(1):69–97.

Alexandre Nikolaev, Yu-Ying Chuang, and R. Harald Baayen. 2022a. A generating model for finnish nominal inflection using distributional semantics. *Mental Lexicon*, 17(3):368 – 394.

Alexandre Nikolaev, Yu-Ying Chuang, and RH Baayen. 2022b. A generating model for finnish nominal inflection using distributional semantics.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Elnaz Shafaei-Bajestan, Masoumeh Moradipour-Tari, Peter Uhrig, and R Harald Baayen. 2022. Semantic properties of English nominal pluralization: Insights from word embeddings. *arXiv*.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Chris Westbury and Geoff Hollis. 2019. Conceptualizing syntactic categories as semantic categories: Unifying part-of-speech identification and semantics using co-occurrence vector averaging. *Behavior Research Methods*, 51(3):1371–1398.

Adam Wiemerslage, Arya D McCarthy, Alexander Erdmann, Garrett Nicolai, Manex Agirrezabal, Miikka Silfverberg, Mans Hulden, and Katharina Kann. 2021. Findings of the sigmorphon 2021 shared task on unsupervised morphological paradigm clustering. In *Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 72–81.

Adam Wiemerslage, Miikka Silfverberg, Changbing Yang, Arya D. McCarthy, Garrett Nicolai, Eliana Colunga, and Katharina Kann. 2022. Morphological processing of low-resource languages: Where we are and what's next.