

Russian Prepositional Phrase Semantic Labelling with Word Embedding-based Classifier ^{*}

Vadim Gudkov¹
vadim0006@gmail.com

Anastasia Golovina ¹
besiage@gmail.com

Olga Mitrofanova ¹
o.mitrofanova@spbu.ru

Victor Zakharov ¹
v.zakharov@spbu.ru

¹Saint Petersburg State University,
Saint Petersburg Russian Federation

Abstract

This paper discusses experiments on automatic extraction, classification and semantic labelling of prepositional phrases in a Russian text corpus. Semantic description of prepositional phrases used in our study is based on G.A. Zolotova's simplified classification of minimal language entities. We present the experimental setup, explain the procedure of the training dataset development, describe the labelling techniques, and provide analysis of results. Our research shows that although semantic differences between some prepositional semantic classes are quite vague, it is possible to achieve promising classification results for core classes.

Keywords: *prepositional phrases, short text classification, word embeddings, semantic labelling.*

1 Introduction

Semantic labelling is understood as the task of annotating language units with labels denoting their semantic meaning or role. Although the most common variation of this task is Semantic Role Labelling, the process of attaching labels carrying semantic role information of the predicate-argument structure to sentence parts, that is not the only way of classifying semantic categories. The experiment described in this paper deals with a wider interpretation of semantic labelling - that is, the process of assigning semantic category labels to syntactic constructions, namely prepositional phrases.

Although there exist multiple theoretical classifications of prepositional phrases in literature with respect to their semantics, and there is strong evidence which suggests that semantic information is highly useful for tasks which traditionally rely on structural features (such as

^{*}Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

parsing), the topic of semantic disambiguation of prepositional phrases is underdeveloped. The study presented in this paper is an attempt to explore this problem.

The preliminary research on the topic of semantic labelling applied to prepositional phrases showed that no classification suitable for NLP tasks had been developed or adapted for the Russian language. Additionally, no corpus big and representative enough specifically for studying prepositional phrases had been available. Therefore, we took the matter into our own hands and created the first corpus of this kind for Russian prepositional phrases. The phrases in our corpus consist of three elements: the head, the preposition and the dependent noun phrase. A more detailed description is offered in Section 3.2.

Having dealt with the corpus issue, we approached the next stage of our task: the selection of the semantic classification to be used for annotating the acquired prepositional phrases. However, the existing semantic interpretations of prepositions and their relations proved to be unfit for use in our study due to their high degree of granularity and the lack of consistency. We ended up adapting one of them to suit the needs dictated by our task. The classification described in this paper is heavily based on the work of our predecessors, however, ours differs in that it was created specifically with the NLP applicability in mind.

In the early stages of our study we took notice of another problem that has to do with the core element of a prepositional phrase - the preposition. Although an indisputable part of the Russian lexicon and morphological system, prepositions remain a class with blurred boundaries. Russian prepositions are commonly divided into non-derivative, simple derivative and complex derivative [Shvedova, 1980]. While the non-derivative prepositions are a closed class, the list of their derivative counterparts remains disputed and varies greatly from source to source, some being considered to be improper prepositions. It is therefore not surprising that although the core, non-derivative prepositions have enjoyed attention in literature, the derivative prepositions are still understudied, especially in computational linguistics.

While generating the prepositional phrase corpus used in this study we paid special attention to derivative prepositions and functional phrases containing prepositions. We went beyond the more commonly observed practice of treating only prepositional phrases with simple prepositions as valid and included complex derivative prepositions into our subject of study as well. We also took heed of functional set phrases containing prepositions as to avoid treating them as part of a prepositional phrase.

The main objective of our experiment was to study the potential of distributed word representations in a vector space model in semantic description of Russian prepositional phrases. In order to do that, we set out to perform a classification of prepositional semantic categories using a supervised machine learning algorithm trained on a corpus of manually tagged prepositional phrases. The primary hypothesis of our study was that word embeddings of prepositional constructions can include representations of semantic categories intrinsic to prepositions. This study was conducted as part of the RFBR project 17-29-09159 “Quantitative grammar of Russian prepositional constructions” carried out at the Department of Mathematical Linguistics, Saint Petersburg State University, Russia [Zakharov, 2017; Zakharov and Azarova, 2019]. The comprehensive goal of the project is to create a representative quantitative lexical-grammatic description of Russian prepositions based on corpus data, with a strong focus on the semantic features of prepositional constructions, as this is something that has not been comprehensively studied as of yet.

The rest of the paper is structured as follows. Section 2 provides theoretical groundings of the prepositional phrase semantic classification applied in our project. In Section 3 we describe our approach and the technical details of our experiment. Section 4 is dedicated to

data analysis as well as to the discussion of tendencies revealed by the output errors. Finally, Section 5 concludes and outlines some directions for future research.

2 Related Work

Traditional linguistic methods are based on the understanding of language as a multilevel system, each linguistic level having its own elementary unit. Modern research practices, however, lean towards an integral view focused on language structures uniting different linguistic elementary units. This view has been developed within Construction grammar (Ch. Fillmore and his followers: [Fillmore, 1988], see also an overview of the trend in [Rakhilina, 2010]). Within this theory, constructions are regarded as complex signs comprising units of various linguistic levels which constitute a functional whole and not a mere sum of their elements. The given approach was adopted in our study. Namely, we assume that the meaning of a preposition defines the semantic features of the whole prepositional phrase, and vice versa, the semantic and syntactic connection between words in a prepositional phrase is understood as a realization of a particular prepositional meaning.

Most of the studies concerned with prepositional phrases have been related to the prepositional phrase attachment disambiguation task as it remains one of the main sources of parsing errors. It has been proven that the use of distributed word representations may improve prepositional phrase attachment accuracy [Agirre et al., 2008; Belinkov et al., 2014; Dasigi et al., 2017]. Therefore, we hypothesize that a similar approach may be helpful in our case as well.

A study akin to ours has been performed and discussed in [Rudzicz and Mokhov, 2010], although the number of classes selected was considerably lower (only 7 classes were chosen).

It must be noted that there exists no commonly accepted semantic classification of Russian prepositional phrases as of yet. As the performance of our system would crucially depend on the semantic classification used, it was necessary to either adapt an existing one to fit the needs of our task or develop our own.

The most obvious approach which implies extraction of prepositional meanings from a dictionary proved to be unfit for the task on hand. Although most dictionaries do offer definitions for functional parts of speech such as prepositions, the word senses they distinguish are not suitable for use in NLP tasks. As such, the three dictionaries we studied, Small Academic Dictionary of the Russian Language, Wiktionary, Explanatory Dictionary of the Russian language by S.I. Ozhegov and N.Ju. Shvedova, had the same critical problems. Firstly, the format of the definitions themselves (their number, volume, content and granularity) fully depend on the format of the dictionary itself. We found the senses used in the dictionaries examined for our task to be too fine-grained, often including rare, orthodox and metaphorical senses. Another problem lay in the lack of any systematic approach to defining the prepositional senses: a preposition is defined with no regard for another preposition's senses, making the creation of a uniform classification entirely impossible. Additionally, no syntactic features of prepositional phrases other than the dependent's case were included in the definitions, making the dictionary senses of prepositions harder to study from the point of view of syntactic units.

A much more appropriate approach was discovered in the Syntactic Dictionary by G.A. Zolotova [Zolotova, 1988]. In this work, G.A. Zolotova introduces the minimal syntactic unit called syntaxeme in order to describe the syntactic structure of the Russian language. A big part of the dictionary is dedicated to prepositional syntaxemes understood as the combination

of a preposition and the case of its nominal dependent, for instance, the preposition ‘в’ plus the accusative case. G.A. Zolotova describes the functional dependency relations of a syntaxeme as well as its semantic meanings, which makes the Syntactic Dictionary a better-suited resource for the purposes of studying semantic features of syntactic constructions. G.A. Zolotova also offers a systematically disjointed yet highly detailed sense classification of syntaxemes, defining such sense types as Location, Direction, Instrument etc., subdivided further in accordance with their dependency relations. Although this classification is also highly fine-grained, we felt it could be successfully adapted for use in our modelling task. Another shortcoming of G.A. Zolotova’s dictionary is the absence of most derivative prepositions, something we also attempted to tackle in our study. An attempt at adapting G.A. Zolotova’s classification for use in NLP tasks can be found in [Mikhailova, 2015]. V.D. Mikhailova revised G.A. Zolotova’s classes, uniting minor senses into bigger ones and giving labels to the senses left un-named in the original classification. However, the reworked classification (further referred to as Zolotova-Mikhailova’s classification) still focused on the semantic features of separate prepositions as opposed to the entire class and proved to be too complex for effective classification.

3 Our Approach

3.1 Generalization of Zolotova-Mikhailova’s classification

In order to reduce the drawbacks of the existing semantic classifications of prepositional senses, we reworked the classes described in Zolotova-Mikhailova’s classification and proposed a single-layer universal classification. To do that, we united the sense subcategories into larger classes while also separating those with double sense labels into distinct classes. We clustered the remaining classes by closeness of sense and united each of them into superclasses, ending up with 15 classes in total.

3.2 Generating a Corpus

In order to prove the hypothesis, we developed a large and representative corpus of Russian prepositional phrases, which at present has no analogues. We used a SynTagRus pretrained UDPipe model [Straka and Straková, 2017] in order to extract the prepositions with the head word and its dependant full noun phrases. The link to the tool developed for the task can be found in the Appendix. The extraction pipeline is displayed in Figure 1.

We used a large dictionary of Russian simple and complex prepositions to verify the parser’s decisions. The preposition list was compiled from dictionaries (Efremova, Sharoff, Rogozhnikova, Small Academic Dictionary, Wiktionary) and corpora (RNC, OpenCorpora, HANCO). Some normalization work was done in order to ensure that all variants of the same preposition would be counted as one, as some prepositions come from the same root and are only different in their form due to their being proclitics, such as ‘о/об/обо’, ‘без/безо’, ‘с/со’ etc. Such variants were therefore concatenated.

Simple and complex prepositions were treated in the same way. Complex prepositions were interpreted as standalone units with their own attachments and dependants.

We also filtered out the prepositional homonyms as well as phrases with prepositions which are parts of functional words and some idiomatic phrases. The list of such stop words was extracted from the Explanatory dictionary of functional parts of speech of the Russian language by T.F. Efremova [Efremova, 2004] and supplemented by the lists of set expressions

Table 1. Generalized Zolotova-Mikhailova's classification

Class	Examples	
Destination	<i>комната для курения</i> (room for smoking)	<i>отправиться на поиски</i> (go on a search)
Direction	<i>плыть против течения</i> (swim against the tide)	<i>подойдет к дому</i> (came up to the house)
Instrument	<i>протрите сквозь сито</i> (push through a sieve)	<i>ловить на удочку</i> (fish with a rod)
Cause	<i>краснеть при мысли</i> (blush at the thought)	<i>плакать из-за ссоры</i> (cry because of a row)
Quality	<i>слишком стар для службы</i> (too old for service)	<i>лучшая среди нас</i> (the best among us)
Quantity	<i>состоять из 3 частей</i> (consist of 3 parts)	<i>вырасти до 1 м</i> (rise up to 1 mil)
Commitment	<i>Татьяна с Ольгой</i> (Tatiana with Olga)	<i>смея сквозь слезы</i> (laughter through tears)
Location	<i>лежит между горами</i> (lies between mountains)	<i>аллея у пруда</i> (path by the pond)
Object	<i>ударить об стол</i> (hit against the table)	<i>сконцентрироваться на работе</i> (focus on the job)
Potential	<i>застраховать от забот</i> (secure from worries)	<i>жадный до денег</i> (hungry for money)
Situation	<i>богослужение под дождем</i> (worship in the rain)	<i>забыли среди суеты</i> (forgot amidst the commotion)
Source	<i>услышала от подружки</i> (heard from a friend)	<i>прочитать в книге</i> (read in a book)
Theme	<i>думал о сыне</i> (was thinking of his son)	<i>консультироваться по вопросам</i> (consult on questions)
Tempus	<i>откроет к ноябрю</i> (will open by November)	<i>воспоминания из детства</i> (memories from childhood)
Transgression	<i>превращение из куколки</i> (metamorphosis from chrysalis)	<i>посадить под арест</i> (put under arrest)

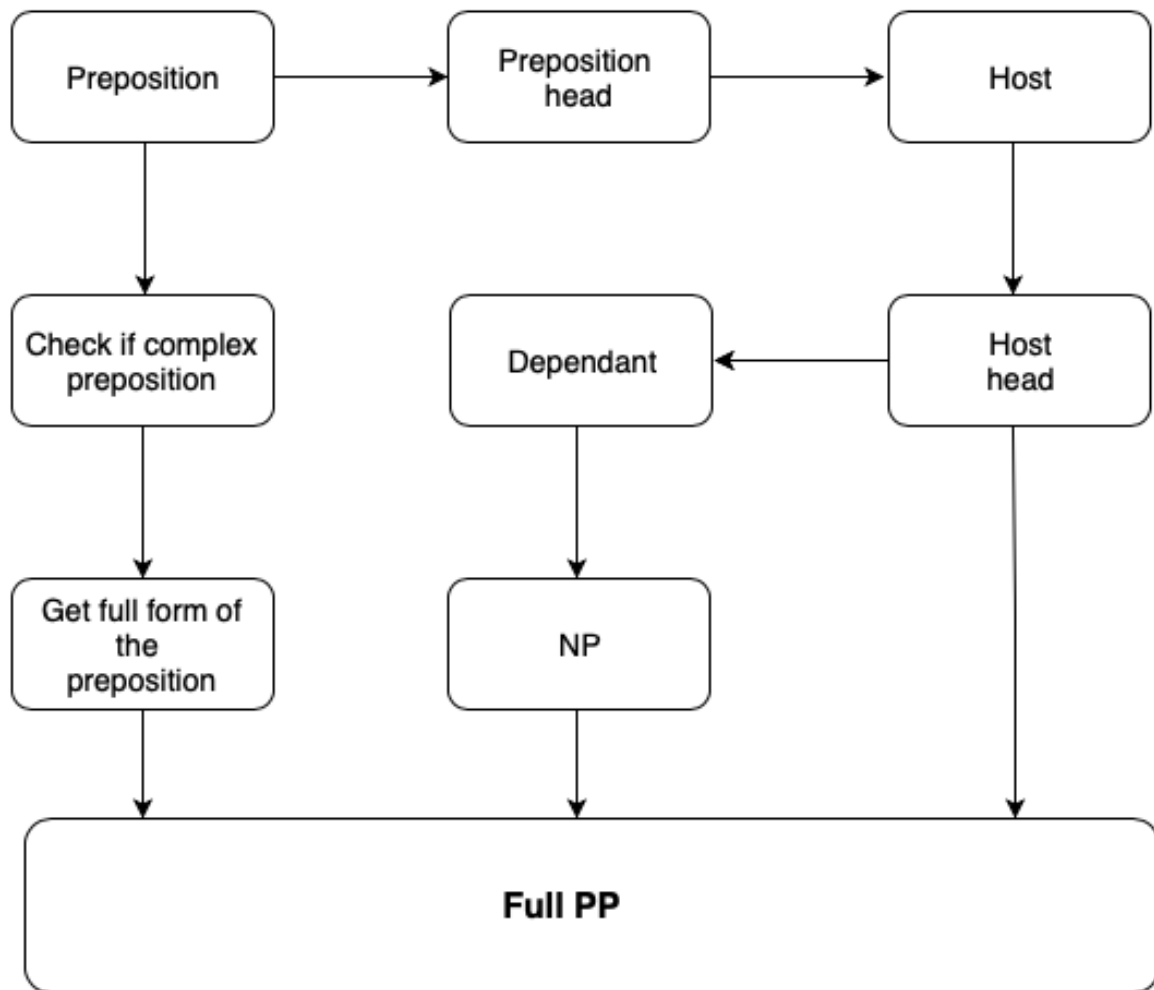


Figure 1: The full cycle of a PP extraction from raw text

containing prepositions provided by the Russian National Corpus. We also added OpenCorpora morphological annotation with the help of PyMorphy2 [Korobov, 2015] for the extracted data in order to ease the subsequent annotation part.

Following this approach we parsed the Taiga news subcorpus [Shavrina, 2018] (size 92 million tokens) and acquired 5 million prepositional phrases. For each prepositional phrase its head, preposition and dependant, alongside with their POS, lemmas and the case of the dependant, were extracted as well. The absolute frequency distribution of the most commonly occurring prepositions observed in the corpus is shown in Figure 2.

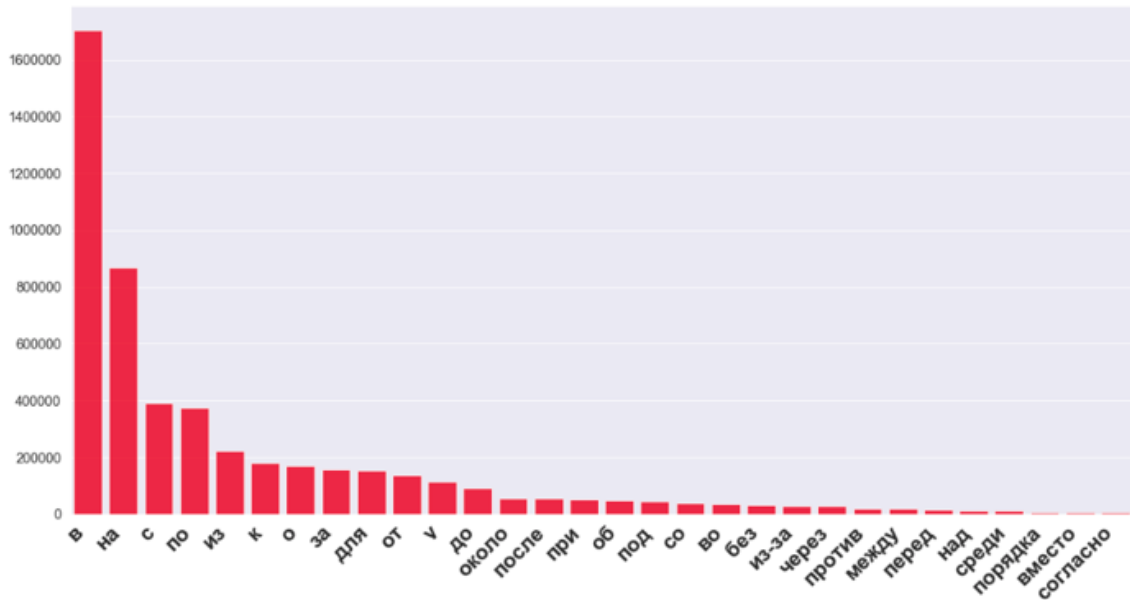
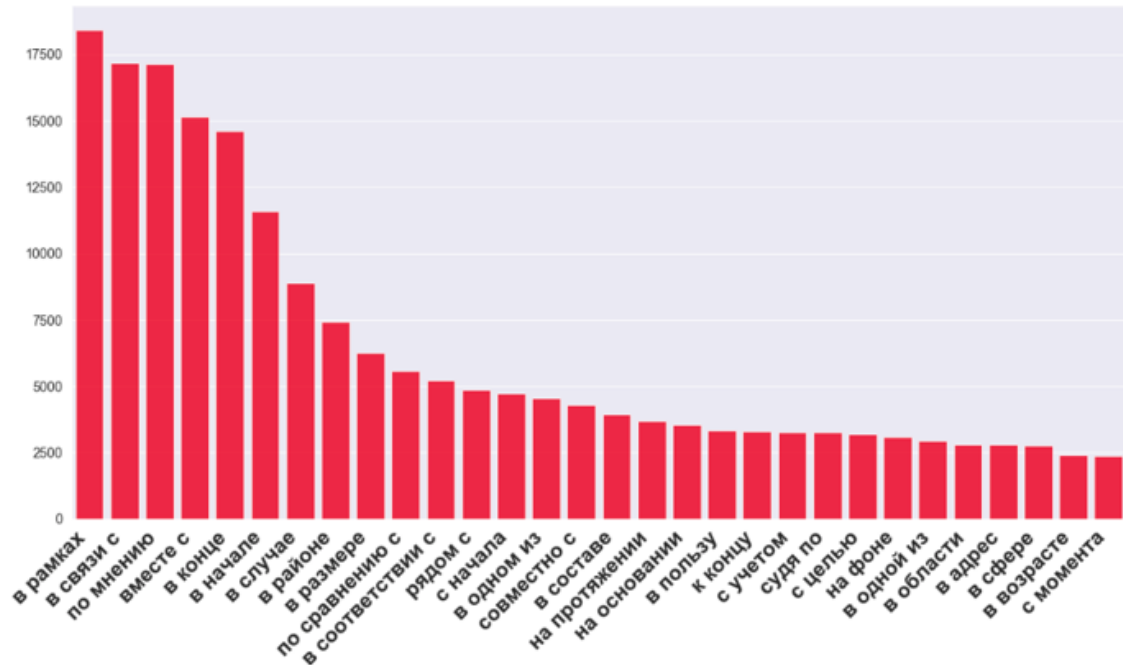


Figure 2: The distribution of 30 most common simple and complex prepositions in the corpus

3.3 Human Annotation

The generalization of Zolotova-Mikhailova’s classification furnished us with the final set of semantic categories characterising prepositional phrases. The classification was manually applied to all of the prepositions described by G.A. Zolotova (mainly non-derivatives), yielding the base set of semantically classified prepositional syntaxemes.

A randomly selected data subset of 10 000 phrases was then acquired from the original 5

million corpus for manual annotation. The previously undescribed prepositional syntaxemes observed in the subset were collected and classified as well.

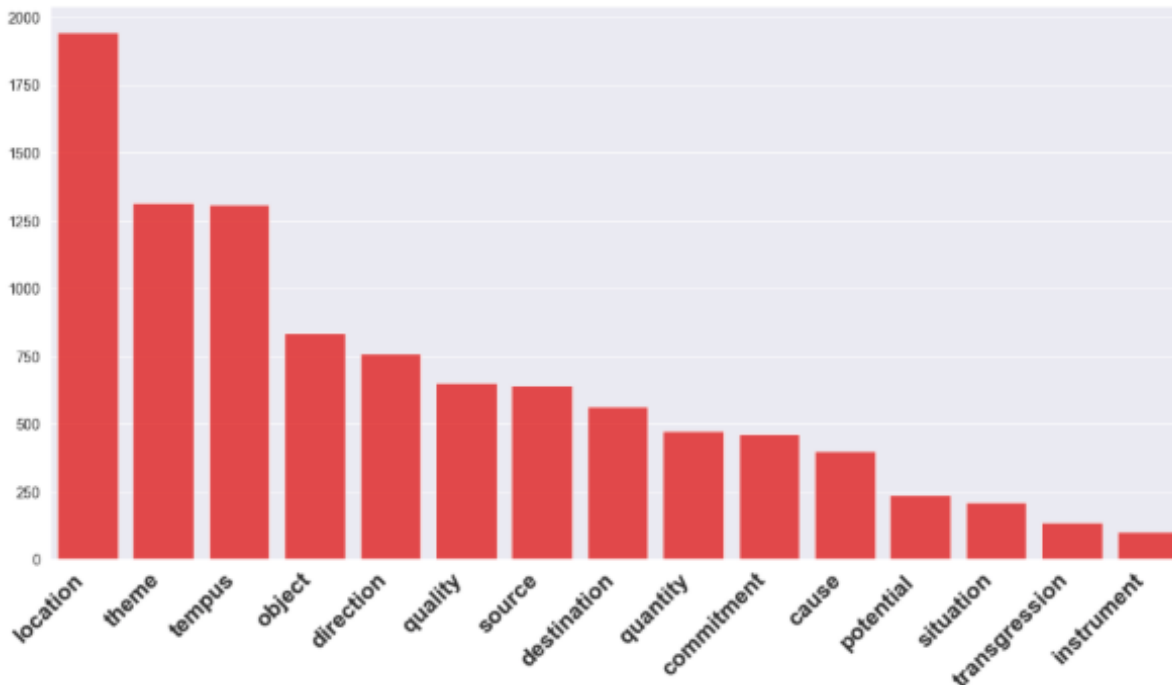


Figure 3: Distribution of semantic classes in the manually labelled subcorpus

After that we created a form containing the list of the syntaxemes, all of the semantic functions of each syntaxeme illustrated with definitions and examples, and tables containing sets of untagged prepositional phrases from the subset sorted by syntaxeme. Thirty annotators with linguistic background were then asked to perform sense disambiguation of prepositional phrases based on the provided classification. After the disambiguation, all of the annotated phrases were manually checked by experts. Figure 3 demonstrates the absolute frequency distribution of the fifteen semantic categories used in our classification.

3.4 Classification Task

After acquiring the labelled dataset we were able to perform a supervised classification. The preliminary analysis of the data uncovered the fact that the semantic classes are highly imbalanced in the corpus. The least represented of them are *Instrument*, *Transgression*, *Situation* and *Potential* as opposed to *Location*, *Theme* and *Tempus*. This imbalance directly affected the performance of the classifier, which leads us to the idea that class enhancement could be an option for further research. However, this experiment is left out of the scope of this paper. In our experiments we revised a standard strategy used in previous works on construction classification [Lyashevskaya et al., 2013]. This strategy develops a view on constructions as multilevel entities combining lexical, semantic, morphological and syntactic features. Context markers constituting a set of constructions for a target word allow the identification of its meaning, so that separate meanings of a polysemous word can be associated with independent clusters of constructions. The procedures of word sense induction/disambiguation and semantic la-

elling thus can be fulfilled by means of super-vised construction classification [Lyashevskaya et al., 2011]. Our approach is based on the assumption that classification of prepositional phrases should be performed not for constructions per se but for their vector representations obtained from Word2Vec models [Mikolov et al., 2013] trained on the Russian corpora from the RusVectores project [Kutuzov et al., 2016]. In this respect, the vector of a prepositional phrase is composed as a sum of vectors corresponding to lexical items constituting the whole construction.

Our preprocessing procedure was performed in accordance with the technique proposed in [Kutuzov and Kuzmenko, 2017], which comprised POS label attachment to the tokens. We made multiple attempts with pretrained embeddings, alt-hough without achieving any viable results with F1 score of 0.4 on average with traditional embedding approaches of TF-IDF and Word2Vec.

We then decided to obtain our own word vectors via FastText supervised [Bojanowski et al., 2016]. This tool enables word embedding via subword information, which has proven to be extremely efficient for morphologically rich languages, such as Russian. Moreover, it utilizes label information via a softmax layer to obtain a probability distribution over predefined classes. With the help of this tool, the classification results were elevated to a higher level. As regards F1 score, the categories *Destination*, *Quantity*, *Location* and *Tempus* are recognized quite effectively, F1 measure being from 0.70 up to 0.86. The other categories achieve moderate F1 values. Due to the highly unbalanced nature of the dataset, some of the lower represented classes were not properly identified. However, average F1 score for the whole set of semantic categories reaches 0.65 and may be boosted after certain improvement of the experimental settings. Data on F1 measure values are given in Table 2.

4 Error Analysis

In order to uncover the tendencies in the classification errors we considered a random subset of 300 prepositional constructions with manually assigned labels and those predicted by a model. We found 95 mismatches between the given label and the predicted one, that roughly corresponds to average F1 score evaluating the classification effectiveness. Thorough analysis of the sample revealed true errors and meaningful mismatches.

However, a major factor influencing correctness of predicted labels was the disbalance in the representation of semantic categories in the training data. As has been mentioned in Section 3.4, several classes were significantly underrepresented in the training dataset. The problem of data sparsity observed for those classes led to their predominance among the mislabelled categories, among which we should mention Potential, Situation, Transgression and Instrument as the least represented categories with the lowest F1 score.

We registered only a small number of true errors exemplified by short contexts with deictic words (e.g. *надо для этого, для этого остаться*, etc.) and/or potentially ambiguous parts (e.g. *стать от такого региона, ставка по программе*, etc.). Set expressions or phraseological units (e.g. *обратиться с просьбой*) defy proper treatment for their possible non-compositionality. In such cases it is hard to obtain the correct label prediction. Nonetheless, in most cases we found meaningful mismatches between labels assigned by experts and those predicted by the model. The most common reason is the inevitable semantic fuzziness of categories included in the generalized Zolotova-Mikhailova’s classification. We found sets of prepositional phrases which should be treated as manifestations of merged semantic classes,

e.g. *Object - Theme* (e.g. *заявлять о задержке зарплаты, сообщать о местонахождении обвиняемого, ходатайствовать о рассмотрении дела, предупредить о закрытии*, etc.); *Tempus - Location* (*находиться в процессе предварительного расследования, видеть в процессе этого занятия*, etc.); *Direction - Location* (*выстрел в воздух, доставить на станцию Акуловка*, etc.); *Destination - Location* (*обращаться во все инстанции*, etc.); *Source - Location* (*прописать в регламенте, отыскаться в решении арбитражного суда*, etc.); *Cause - Quality* (*откликнуться на предложение*, etc.), and so on. In the aforementioned cases the differences between the semantic categories are subtle, so that prepositional phrases with dual labels seem to be immanently ambiguous as they possibly reveal both categories simultaneously. Such prepositional phrases can obtain proper treatment as constructions with diffuse meaning according to Ju.D. Apresjan [Apresjan, 1971]. In such cases differentiation of close meanings turns out to be impossible.

5 Conclusion and Future Work

In this paper we proposed a challenging solution to the problem of automatic semantic classification of prepositional phrases. Although some attempts into classifying the semantic functions of prepositions had been made, none proved to be effective from the point of view of their application in NLP tasks. We used G.A. Zolotova’s system of syntaxeme senses to develop a semantic classification of prepositional phrases which could be used in our task. We then tested the developed classification in manual prepositional phrase sense disambiguation and used the obtained data for training a word embedding-based classifier. The results of the subsequent automatic classification of prepositional phrases presented in this article show that some of the specified classes, such as *Location*, *Tempus*, *Quantity* and *Destination*, can be identified with a high degree of success. At the same time, a few classes underrepresented in the training data and having broader definitions presented a problem to the classifier.

The results attained with our method suggest several possible directions for further development. The problem of data sparsity that affected the predictions received for the smaller classes might be resolved through WordNet-based class enhancement, as was mentioned in Section 3.3. Alternatively, the issue could be overcome by means of reworking the classification itself to assign the prepositional phrases currently found in the poorer-performing classes to the others. In addition to the aforementioned lines of practical research, the proposed classification of prepositional syntaxemes could be used for describing the previously unexplored derivative prepositions. All in all, being the first of its kind for Russian prepositional phrases, our study offers substantial food for thought for further research and experiments.

6 Appendix

The tool used for prepositional phrase extraction and semantic classification, and the data are available at <https://github.com/merionum/ppphrase>.

Acknowledgements

Our research was supported by the RFBR grant 17-29-09159 “Quantitative grammar of Russian prepositional constructions” (2018-2020).

Authors wish to express their sincere gratitude to A.Ts. Masevich, U.V. Butorova, E.G. Filimonov, D.A. Alfimova, A.D. Ivanova, the students of the SPbU Mathematical Linguistics Department and other participants for their valuable help in annotating the data.

References

- [Shvedova, 1980] Shvedova N.Ju. (1980) Russian Grammar. Vol. 1. Moscow, Nauka: 1980. – 792 p. (in Rus.) = Russkaja grammatika. Tom 1. Moscow, Nauka: 1980. – 792 s.
- [Zakharov, 2017] Zakharov, V.P. (2017). A construction grammar approach to Russian prepositions. SGEM 4th International Multidisciplinary Scientific Conference on Social Sciences and Arts 20172367-5659, 2(3), pp. 279-286.
- [Zakharov and Azarova, 2019] Zakharov, V.P., and Azarova, I.V. (2019). Towards a computational ontology of Russian prepositions. Trudy mezhdunarodnoj konferencii “Korpusnaja lingvistika”, pp. 155-165.
- [Fillmore, 1988] Fillmore, C.J. (1988). The Mechanisms of “Construction Grammar”. Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society (1988), pp. 35-55.
- [Rakhilina., 2010] Rakhilina, E.V. (2010) Construction linguistics. Moscow, Azbukovnik Publ.: 2010. – 584 p. (in Rus.) = Lingvistika konstrukcij. Moscow, Azbukovnik: 2010. – 584 s.
- [Agirre et al., 2008] Agirre, E., Baldwin, T., Martínez, D. (2008). Improving Parsing and PP Attachment Performance with Sense Information. In Proc. of the 46th Annual Meeting of the Association for Computational Linguistics, pp. 317–325.
- [Belinkov et al., 2014] Belinkov, Y., Lei, T., Barzilay, R., Globerson, A. (2014) Exploring Compositional Architectures and Word Vector Representations for Prepositional Phrase Attachment. Transactions of the Association for Computational Linguistics, 2, pp. 561-572.
- [Dasigi et al., 2017] Dasigi, P., Ammar, W., Dyer, C., Hovy, E. (2017) Ontology-Aware Token Embeddings for Prepositional Phrase Attachment. In Proc. of the 55th Annual Meeting of the Association for Computational Linguistics (Vol 1: Long Papers), pp. 2089–2098.
- [Rudzicz and Mokhov, 2010] Rudzicz, F., and Mokhov, S. (2010) Towards a Heuristic Categorization of Prepositional Phrases in English with WordNet. Technical Report, Cornell University, (2003), arxiv1.library.cornell.edu/abs/1002.1095-?context=cs
- [Zolotova, 1988] Zolotova G.A. (1988) Syntactic dictionary: Repertory of elementary units of Russian Syntax. Moscow: Nauka, 1988. – 440 p. (in Rus.) = Sintaksicheskij slovar’: repertuar elementarnykh jedinic russkogo jazyka. Moscow: Nauka, 1988. – 440 s.
- [Mikhailova, 2015] Mikhailova, V.D. (2015) Ontology of Prepositions in the Russian language. Saint Petersburg, SPbU: 2015. - 159 p. (in Rus.) = Ontologija predlogov v russkom jazyke. Saint Petersburg, SPbU: 2015. - 159 s. [Straka and Straková, 2017] Milan S., and Straková, J. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, Vancouver, Canada, August 2017.

- [Efremova, 2004] Efremova T.F. Explanatory dictionary of functional parts of speech of the Russian language. Moscow: Astrel, AST, 2004. - 815 p. (in Rus.) = Tolkovyj slovar' sluzhebnykh chastej rechi. Moscow, Astrel: 2004. – 815 s.
- [Korobov, 2015] Korobov M. (2015) Morphological Analyzer and Generator for Russian and Ukrainian Languages // Analysis of Images, Social Networks and Texts, pp 320-332. [Shavrina, 2018] Shavrina, T.O. (2018) Differential Approach to Web-Corpus Construction. In Proc. of the 2018 Annual International Conference “Dialog”. Moscow, 2018. 11 p.
- [Lyashevskaya et al., 2013] Lyashevskaya O.N., Panicheva P.V., Mitrofanova O.A. (2013) Data visualization for building the catalogue of Russian lexical constructions (based on RNC). In: Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue” (2013). Issue 12. Vol. 1 P. 465-478.
- [Lyashevskaya et al., 2011] Lyashevskaya O., Mitrofanova O., Grachkova M., Romanov S., Shimorina A., and Shurygina A. (2011) Automatic Word Sense Disambiguation and Construction Identification Based on Corpus Multilevel Annotation. In: Text, Speech and Dialogue. Proceedings of the 14th International Conference TSD 2011, Pilsen, Czech Republic, September 1–5, 2011. Springer-Verlag, 2011.
- [Mikolov et al., 2013] Mikolov et al. (2013) Distributed Representations of Words and Phrases and their Compositionality. In: Proceedings of Neural Information Processing Systems. 26 p. [Kutuzov and Kuzmenko, 2017] Kutuzov A., and Kuzmenko E. (2017) WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models. In: Ignatov D. et al. (eds) Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science, vol 661. Springer, Cham.
- [Bojanowski et al., 2016] Bojanowski, P., Grave, E., Joulin, A., Mikolov, T. (2016) Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics.
- [Apresjan, 1971] Apresjan Ju.D. (1971) On regular polysemy. In: Proceedings of the Academy of Sciences of the USSR. Department of Literature and Language. Vol. XXX. Issue 6. - M., 1971. - P. 509-523. (in Rus.) = O reguljarnoj mnogoznachnosti. Izvestija AN SSSR. Vol. XXX. Vyp. 6. - M., 1971. - s. 509-523.