Problems of Disambiguation of Prepositional Phrases

Kirill Boyarsky[a,b], Eugeny Kanevsky[b], and Anastasia Kozlova[a]

[a] *ITMO University, Kronverkskiy Ave, 49-A, St. Petersburg, 197101, Russia*
[b] *Institute of Regional Economics Problems RAS, Serpukhovskaya St, 38,* St. Petersburg, *190013*, Russia

Abstract

This paper describes the features that appear in parsing procession of multiword turns (phrasemes) able to act as prepositions. These features are considered in the context of automatic analysis of Russian texts. Such phrases have a fairly high homonymy, which creates some difficulties in analysis and defining semantics and, consequently, reduces the accuracy of parsing. More than 320 phrasemes have been classified on the basis of the assumed homonymy types.

In the course of the study, the phrasemes have been divided into three groups. The first group includes those phrasemes that can definitely be called prepositions, but potentially have some semantic ambiguity. The second group combines phrasemes that are characterized by the part-of-speech homonymy of preposition/adverb. The third group is characterized by phrasemes that determine the construction of two or three parsing options. The occurrence of multivariate parsing is based on the presence of one or two phrases related to different parts of speech, and a simple conjunction of a preposition with a noun.

Within each group, lists of the most common phrasemes have been composed (according to the NCRL), indicating the probability that a certain phraseme may serve as a preposition. The paper also defines the basis on which the compilation of effectively removing homonymy rules for the SemSin parser may rely on. The examples provided in this paper prove that it is necessary to consider not only the direct encirclement of the phraseme, but also its remote context to remove homonymy.

Keywords

automatic text analysis, disambiguation, homonymy, idiomaticity, prepositional phrases

1. Introduction

In the process of automatic parsing of the Russian language sentences and building a dependency tree, there is an arising problem of removing homonymy of various types – morphological, lexical, part-of-speech, etc. One of the ways to solve this problem is the broad use of standard combinations of words – phrasemes. This term refers to a wide range of expressions with a varying degree of idiomaticity [1]. The common feature for phrasemes is that the value of the whole is not a composition of the values of the constituent parts. In general, the words that are part of phrasemes can change, however within the scope of this study we are interested in invariable phrasemes, most of which are turns of speech that perform the functions of:

- adverbs – без царя в голове (‘one who has bats in the belfry’), без конца и края (‘stretching boundlessly’), без устали (‘tirelessly’), …;
- prepositions – без согласия (‘without consent’), в память о (‘in memory of”), за неимением (‘for lack of’ or ‘failing’), на пути к (‘on the way to’), …;
- inserted clauses – а может быть (‘and maybe’), в лучшем случае (‘at best’), видишь ли (‘you see’), …;
- conjunctions – а вместе с тем (‘and at the same time’), в связи с чем (‘in connection with what’), разве только (‘unless’), …;

IMS 2021 - International Conference "Internet and Modern Society", June 24-26, 2021, St. Petersburg, Russia
EMAIL: boyarin9@yandex.ru (A. 1); eak300@mail.ru (A. 2); stasia.kozlova@gmail.com (A. 3)
ORCID: 0000-0002-0306-8276 (A. 1); 0000-0002-1498-4632 (A. 2)

- particles – а то что ж (‘and then what’), едва ли не (‘almost’), как бы (‘as if’), чуть было не (‘nearly’), …;
- predicative turns of speech – пруд пруди (‘a dime a dozen’), раз плюнуть (‘not a big deal’).

The most complete lists of turns of speech are given in the NCRL (National Corpora of Russian Language) [2]. The dictionaries of Kuznetsov [3] and Rogozhnikova [4] have also been used.

Currently, close attention is drawn to the semantics of prepositional groups, including those where more-than-one-word combinations act as a preposition [5]. As even a preliminary analysis shows, most of these phrasemes do not have homonymy and are always prepositional turns.

However, it is possible that the same combination of several words can correspond to two different turns. For example, a phraseme *без сопровождения* (‘unaccompanied’) can function as either a preposition or an adverb, depending on the context of the word on the right: a word in the genitive case, verb, or punctuation mark:[2]

● *Птенцы могли лететь через океан* ***без сопровождения*** *родителей* (‘The fledglings could fly across the ocean without their parents accompanying them’).

● *Медленно,* ***без сопровождения*** *запел хор* (‘Slowly, unaccompanied, the choir began to sing’).

● *Если ребенок выезжает* ***без сопровождения***, *он должен иметь при себе кроме паспорта нотариально оформленное согласие…* (‘If the child goes unaccompanied, he must have with him, in addition to the passport, a notarized consent’).

A more complex situation arises in the event that a combination of several words, depending on the context, may or may not be a turn. Many word combinations of this kind are considered by Rogozhnikova [4], who notes the possibility of their use as free phrases that are homonymous to turns. So, for example, a phraseme *с целью* (‘for the purpose of’) can either perform the functions of a preposition or remain a free word combination, depending on the presence or absence of a word on the right in the genitive case:

● *Испания требовала экстрадиции, выдвигая также обвинения в нарушении прав человека, массовых пытках и заговоре* ***с целью*** *пыток* (‘Spain has demanded extradition, charging accusations of human rights violations, mass torture and conspiracy to torture’).

● В августе, вероятно, с целью отвлечь население от дум о хлебе насущном, было объявлено о создании Комитета по чрезвычайному положению (‘In August, probably in order to distract the population from thinking about their daily bread, the creation of a State of Emergency Committee was announced’).

An even more complex situation is possible, when the same phraseme can serve as a preposition, adverb, or remain a free phrase, depending on the type of right context:

● *Поэтому я пошел* ***по пути*** *референтных групп* (‘So I went the way of reference groups’).

● *Мы ехали на концерт и* ***по пути*** *притормозили на Садовой, у дома Булгакова* (‘We were on our way to a concert and stopped on Sadovaya Street, near Bulgakov's house.’).

● *По пути в Женеву Леня сел за руль* (‘On the way to Geneva, Lenya took the wheel’).

Taking into account the above, all prepositional turns (and the corresponding phrasemes), depending on their structure and the method of analysis used in the parser, in our opinion, can be divided into three groups, which are to be considered below.

There are two approaches to analyzing such turns. The first approach does not involve any special graphematic separation of them – an example of it is the "ETAP-3" parser [6]. In the context of the second approach, such a turn is emphasized in a special way – an example of it is the ABBYY parser [7]. Recently, these approaches are converging, and in the latest version of ETAP-4 [8], some of the turns are also emphasized (combined into a single token). The principle of operation of our SemSin parser [9], which analyzes prepositional turns, is close to the second example. More-than-one-word phrasemes are combined into a single token [10].

It should be noted that the SemSin parser is designed for analyzing written Russian-language texts, mainly newspaper and scientific profiles. The parser consists of 4 blocks: a dictionary, a morphological analyzer, production rules, and a lexical analyzer. The regular paragraph of the Russian-language text undergoes the morphological analysis with the marking out of individual

[2] Here and further on, all the examples are taken from the NCRL and are separated by a "●" sign, and Russian-language phrasemes that are turns of speech are highlighted in bold in the examples. The words that allow to make a particular decision are underlined in Russian-language examples.

tokens (words, phrases, punctuation marks, numbers, etc.). The token chain is then processed in the lexical analyzer using a system of production rules, the purpose of which is to transform the linear sequence of tokens into a dependency tree.

The principles of building the parser dictionary are based on the ideas of Tuzov [11] The main table of the dictionary contains more than 195 thousand lexemes distributed over 1700 classes [12]. Each lexeme has morphological characteristics, as well as the number of its semantic class and actants or valences (for connecting dependent words) in the form of cases (!Nom, !Gen, !Acc, etc.) or prepositions, possibly with the corresponding cases (!Without, !For, !inAcc, !onPrep, etc.). Free actants are also used, which define more generalized concepts (!Question, !Where, !How, !Fromwhere, !Why, etc.). Often, before such an actant, the acceptable classes of words that can replace them are indicated. The presence of a classifier can significantly reduce ambiguity and is especially widely used when connecting adjectives and prepositions. About 14% of words in the dictionary have two or more lexemes.

In addition to the main table, there are auxiliary tables that provide the execution of tasks that are of interest in this work. This is a table of word combinations (more than 5350 lines), containing stable combinations of words with different types of inflection. These can be collocations (*вид на жительство*, 'residence permits'), names of organizations (*Чейз Манхеттен Банк*, 'Chase Manhattan Bank'), or idiomatic expressions (*белая ворона*, 'sore thumb'). In these cases, one or all of the words can be used in different word forms.

In this paper, we are interested in immutable phrasemes that form compound prepositions, adverbs, etc. If the parser decides that a certain phrase is such a phraseme, then the words included in it are combined into a single token.

The second auxiliary table is a table of prepositions (more than 2460 lines) with the cases and semantic classes of the connected nouns. If the connection of the preposition with the dependent word is syntactic in nature and, as a rule, coincides with the case of the dependent word, then the connection of the prepositional group to the main word reflects the semantics more fully (Where, When, Why, etc.).

2. Group 1. Prepositional phrases without lexical homonymy

Passing on to the analysis of prepositional phrases, we note that the largest of the three groups is the first one, which contains turns of speech that have almost no homonyms and are unambiguous[1].The analysis of these prepositions does not differ from the analysis of ordinary one-word prepositions. The group consists of two subgroups: the phrasemes of the first end with nouns (1A), the second – with prepositions (1B).

2.1. Subgroup 1A

This subgroup includes unambiguous prepositional phrases, whose phrasemes consist of two words: a preposition and a noun. There are more than 60 such phrases in our dictionary.

The vast majority of them require the genitive case after them.

Example: ***В конце*** *своего пребывания школьники защищают исследовательские работы на конференции,* ***в присутствии*** *всего коллектива Центра* ('At the end of their stay, students defend their research papers at a conference, in the presence of the entire staff of the Centre').

Other turns require the dative case after them: *в противовес* ('in contrast to'), *в противоположность* ('contrary to'), *в ущерб* ('to the detriment of'), *на благо* ('for the benefit of'), *на радость* ('to smb's joy').

Example: *Отдавать все силы организации выборов* ***в ущерб*** *профессиональной деятельности* ('Give all the effort to organize elections to the detriment of professional activity').

Most prepositional phrases of this type can connect to the main word with only one connection. However, there are also such phrases with semantic homonymy that have two connections for connecting to the main word, the choice of one of which depends on the main word (its class, its internal actants, or in general its part of speech). The following turns refer to this type: *в честь* ('in

honour of’) (What for, Why), *из числа* (‘from the number of’) (From, Which), *на основе* (‘on the basis of’) (How, Which), *по поводу* (‘concerning’) (Dat, Why).

The semantics of prepositional relations is examined in detail in the dictionary of Zolotova [13], but there is a lack of formal rules that allow to correlate mainly syntactic relations developed by the parser with the semantics of [13]. This is a rather complex task that is still under consideration in some special cases [14].

Below are the examples of two prepositional phrases, and the connection with the main word of prepositional phrase is presented in parentheses in terms of the SemSin parser and in the semantic connections of Zolotova.

● *В прессе отмечалось, что это был салют в 101 залп* ***вчесть*** *возникшего в России рабочего вопроса* (‘In the press it was noted that it had been a salute of 101 volleys in honor of the labour issue that arose in Russia’) –(был – Зачем (финитив) – в честь) (Why (finitive)).

● *И вдруг Олег вспомнил, как однажды он был на торжественном ужине, устроенном* ***в честь*** *приезда английского принца Чарльза* (‘And suddenly Oleg remembered how he once was at a gala dinner, arranged in honour of the arrival of English Prince Charles’) – (устроенном – Почему (каузатив) – в честь) (Why (causative)).

● *Если опальный магнат будет исключён* ***из числа*** *сопредседателей ЛР, у партии возникнут финансовые затруднения, полагают влиятельные эксперты* (‘If the disgraced magnate is excluded from the number of co-chairs of the Republic of Latvia, the party will face financial difficulties, influential experts believe’) – (исключён – Изо (финитивно-фазисное) – из числа) (From (phase-finitive)).

● *Обычно в то время, как наверху происходила церемония награждения победителей, внизу совершалась казнь изменников, трусов, неудачников* ***из числа*** *подданных Великого курфюрста* (‘Usually, while the ceremony of awarding the winners took place above, the execution of traitors, cowards, losers from among the subjects of the Prince-elector was carried out below’) (неудачников– Какой (генератив) – из числа) (Which (generative)).

Table 1 shows the most common prepositional phrases of this subgroup that require the genitive case after them. A questionof the validity of this table arises. Obviously, expert evaluation is very difficult in this case because of the necessity to view too many sentences. For example, for a phraseme ***в глубь*** (‘into the depth’), it would be necessary to analyse more than 27 hundred sentences in order to identify about 70 cases of absence of a word in the genitive case to the right of the phraseme. It is very likely that if we choose 300-500 sentences in any way, there will be no cases of absence of the genitive case on the right.

Therefore, such method of evaluation has been chosen. With the usage of the capabilities of the NCRL, sentences in which there is a punctuation mark after the studied phraseme *в глубь* (‘into the depth’) have been selected.

Example: *И светлый месяц, который то серебрил всё море, рассыпая по мелкой ряби свои лучи, то одним цельным блиставшим столбом падал* ***в глубь****, перерезав всю бухту* (‘And the bright moon, which now silvered the whole sea, scattering its rays on the faint ripples, then fell in one solid shining column into the depths, cutting the entire bay’).

It is obvious that in all these sentences (171 units) this phraseme does not serve as a preposition, but is simply a combination of a noun with a preposition. Next, sentences, in which there is a verb in the indicative or imperative mood, an infinitive or an adverbial participle after the studied phraseme, have been selected. This set of 57 sentences requires expert analysis, since after this phraseme there are such homonymous words as *души* (‘souls’ vs ‘strangle’), *заросли* (‘thickets’ vs ‘overgrow’), *моря* (‘seas’ vs ‘starve’), *села* (‘villages’ vs ‘sit’), *стекла* (‘glasses’ vs ‘drain’), *суши* (‘land’ vs ‘sushi’ vs ‘dry’), etc.

Example: *Мы уложили вещи и двинулись* ***в глубь*** *села* (‘We packed our bags and moved to the heart of the village’).

In only 6 sentences, our phraseme is simply a combination of a noun and a preposition.

The sum of these sentences (65+6) determines the reliability of the fact that this phraseme serves as a preposition.

Table 1.
Prepositions that require the genitive case

Turn of speech	Link with the main word	Frequency, ipm	From which preposition
В ВИДЕ (‘in the form of’)	Как (‘How’)	35.5	99.4%
В ГЛУБЬ (‘into the depth’)	Куда (‘Where’)	8.4	97.5%
В ПОЛЬЗУ (‘in favour of’)	Как (‘How’)	15.2	97.2%
В ПРИСУТСТВИИ (‘in smb’s presence’)	Как (‘How’)	20.5	99.2%
В ТЕЧЕНИЕ (‘during’)	какДолго (‘For how long’)	57.8	99.2%
В ХОДЕ (‘in the course of’)	Когда (‘When’)	18.6	99.8%
В ЦЕЛЯХ (‘with a view to’)	Для (‘For’)	8.2	97.6%
В ЧЕСТЬ (‘in honour of’)	Зачем, Почему (‘For what reason’, ‘Why’)	7.9	98.1%
ВО ВРЕМЯ (‘during’)	Когда (‘When’)	212.7	99.2%
ВО ИМЯ (‘in the name of’)	Зачем (‘For what reason’)	13.1	99.2%
ДЛЯ СОЗДАНИЯ (‘for creating’)	Зачем (‘For what reason’)	6.8	99.1%
ЗА ПРЕДЕЛЫ (‘outside the limits of’)	Куда (‘Where’)	12.3	98.0%
ЗА СЧЕТ (‘at the expense of’)	Как (‘How’)	27.3	99.6%
ИЗ ЧИСЛА (‘from the number of’)	Изо, Какой (‘From’, ‘What’)	11.8	99.7%
НА ОСНОВАНИИ (‘on the grounds of’)	Почему (‘Why’)	31.8	99.3%
НА ОСНОВЕ (‘on the basis of’)	Как, Какой (‘How’, ‘What’)	23.5	99.6%
НА ПРОТЯЖЕНИИ (‘throughout’)	какДолго (‘For how long’)	12.2	99.8%
ПО ПОВОДУ (‘concerning’)	поДат, Почему (‘Dative’, ‘Why’)	30.3	99.2%

2.2. Subgroup 1B

This subgroup includes unambiguous prepositional phrases, whose phrasemes consist of two, three or four words and end with a preposition. There are more than 150 such phrases in our dictionary. Almost all prepositional phrases ending with a preposition belong to this subgroup. To date, we know only three exceptions: phrasemes *на глазах у* (‘before smb's eyes’), *под носом у* (‘under the nose of’), and *под самым носом у* (‘under the very nose of’). Indeed, let us compare two sentences: *Ты* ***на глазах у*** *зрителя вершишь свой путь* (‘You are making your way before the eyes of the viewer’) and *Стали мы во дворе, и вижу я: на глазах у него будто слеза поблескивает* (‘We are standing in the courtyard, and I see: in his eyes, a tear seems to glisten’). It is quite obvious that in the first sentence the phraseme is a prepositional phrase, while in the second one it is just a free combination of three words. The situation is similar with the other two phrasemes. All of them belong to the third group.

Examples of the most frequent turns of speech of subgroup 1B are given in Table 2. As the table shows, most of them begin with a preposition, usually it is «в» (‘in’). At the end, the prepositions «с» «со» (‘with’) or «от» (‘from’) are most often located. The case required after the turn is determined by the preposition in the end. Due to the presence of a preposition in the end, the question of the

reliability of the data does not arise – theoretically, an adjective, a participle, a pronoun or in the appropriate case should always be to the right of the preposition (otherwise it is just an error in the text).

Table 2.

Phrasemes ending with a preposition

Turn of speech	Required case	Link with the main word	Frequency, ipm
В ОДНОЙ ИЗ (‘inoneof’)	Род (‘Genitive’)	Где (‘Where’)	17.8
В ОДНОМ ИЗ (‘inoneof’)	Род (‘Genitive’)	Где (‘Where’)	28.8
В ОТВЕТ НА (‘in response to’)	Вин (‘Accusative’)	Как (‘How’)	18.1
В ОТЛИЧИЕ ОТ (‘incontrastto’)	Род (‘Genitive’)	Как (‘How’)	21.4
В СВЯЗИ С (‘in connection with’)	Тв (‘Instrumental’)	Почему (‘Why’)	35.3
В СООТВЕТСТВИИ С (‘inaccordancewith’)	Тв (‘Instrumental’)	Как (‘How’)	31.6
ВМЕСТЕ С (‘together with’)	Тв (‘Instrumental’)	Как (‘How’)	89.9
ВМЕСТЕ СО (‘together with’)	Тв (‘Instrumental’)	Как (‘How’)	27.1
ВНЕ ЗАВИСИМОСТИ ОТ (‘regardlessof’)	Род (‘Genitive’)	Как (‘How’)	32.1
ВПЛОТЬ ДО (‘up to’)	Род (‘Genitive’)	Как (‘How’), доКогда, Докуда, Сколько	27.6
ВСЛЕД ЗА (‘following’)	Тв (‘Instrumental’)	Как (‘How’)	21.4
НЕ БЕЗ (‘not without’)	Род (‘Genitive’)	сТв (‘With’)	38.1
НЕСМОТРЯ НА (‘in spite of’)	Род (‘Genitive’)	Как (‘How’)	45.6
ПО НАПРАВЛЕНИЮ К (‘in the direction of’)	Вин (‘’)	Куда (‘Where’)	86.4
ПО ОТНОШЕНИЮ К (‘inrelationto’)	Дат (‘Dative’)	поОтн (‘inrelationto’)	40.9
ПО СРАВНЕНИЮ С (‘in comparison with’)	Дат (‘Dative’)	Как (‘How’)	20.3
РЯДОМ С (‘near to’)	Тв (‘Instrumental’)	Где (‘Where’)	55.6
ЧТО ДО (‘as for’)	Тв (‘Instrumental’)	Как (‘How’)	22.3

Most prepositional phrases of this type can connectto the main word in only one link. However, there are also such phrases that have several links to the main word, the choice of one of which depends on the main word (its class, its internal actants, or in general its part of speech). The following turns refer to this type: *верхом на* (‘astride’) (How, To where), *вплоть до* (‘up to’) (How, How long, How far, How much), *начиная от* (‘starting from’) (How, When), *начиная с* (‘starting from’) (How, When), *начиная со* (‘starting from’) (How, When), *совместно с* (‘together with’) (How,Instr), *совместно со* (‘together with’) (How, Instr).

Below the examples for the prepositional phrase *вплоть до* (‘up to’) are given.

● *Помощь готова оказать любую,* ***вплоть до*** *аврального написания сочинения* (‘I am ready to provide any help, up to the emergency writing of an essay’) – (*оказать*– Как (Интенсив) – *вплоть до*) (How (Intensive)).

● ***Вплоть д****о 1933 года прокуратура входила в состав Народного комиссариата юстиции* (‘Until 1933, the Prosecutor's Office was part of the People's Commissariat of Justice’) – *(входила – доКогда (Темпоратив) – вплоть до)* (How long (Temporative)).

● *То развенчание "культа личности", то внедрение кукурузы* ***вплоть до*** *Полярного круга, то построение коммунизма в одной отдельно взятой стране*…(‘The debunking of the "cult of personality", the adoption of corn up to the Arctic Circle, the construction of communism in one single country…’)– (*внедрение* – Докуда (Директив) – *вплоть до*) (How far (Directive)).

● *С помощью частиц, разогнанных на ускорителях, мы можем сегодня зондировать расстояния* ***в плоть до*** 10^{-16} (‘With the help of accelerated particles, we can now probe distances up to 10^{-16}’)(*зондировать*– Сколько (Дименсив-квантитатив) – *вплоть до*) (Howmuch (Dimensive-quantifier)).

3. Group 2. Phrases with the preposition/adverb homonymy.

This group includes the simplest homonymous prepositional phrases, whose phrasemes can serve as prepositions or adverbs [15]. In our dictionary, there are more than 20 such turns. For example, a phrase *накраю*(‘on the verge’) can be a preposition if it is followed by a word in the genitive case, or an adverb in case of its absence:

● *Они остановились* ***на краю*** *заполненного серым туманом гигантского провала* (‘They stopped at the edge of a giant chasm filled with gray fog’).

● *Если бы, Саша, ты успел еще что-нибудь во славу русского национализма высказать, носить бы нам тебе передачки, а так как-то удержался на краю...*(‘If you had had time to say anything else to the glory of Russian nationalism, Sasha, we would have had to bring you parcels, but somehow you stayed on the edge...’)

The vast majority of prepositional turns require the genitive case after them. Example:

● *Родиться князем не мудрено, и можно* ***по праву*** *породы называться сиятельством*.

Two phrases require a dative case after them: *в угоду* (‘to please’), *не в пример* (‘unlike’). Example:

● *Он просто не хотел никого казнить* ***в угоду*** *иудеям* (‘He just didn't want to execute anyone to please the Jews’).

Most prepositional phrases of this type can connectto the main word in only one link. However, there are several turns that have two connections for connecting to the host, the choice of one of which depends on the main word (its class, its internal actants, or in generalits part of speech). The following turns refer to this type: *в конце*(‘at the end’) (Where, When), *в начале*(‘at the beginning’) (Where, When), *в середине*(‘in the middle’) (Where, When), *к концу*(‘to the end’) (When, To where), *к началу*(‘to the beginning’) (When, To where). Below are examples for the prepositional phrase *в начале* (‘at the beginning’).

● *Я только успел заметить далеко* ***в начале*** *улицы две светлых фигурки* (‘I only had time to notice two light figures far away at the beginning of the street’) – (заметить – Где (Локатив) – в начале) (Where (Locative)).

● *Да, а* ***в начале*** *марта мы-таки устроим массовый вылет* (‘Yes, and in early March, we will still arrange a mass flight’) (устроим – Когда (Темпоратив) – в начале) (When (Temporative)).

When the phrasemes of the second group are detected, the parser also combines the words included in them into a single token, but outputs two lexemes that are present in the dictionary: a preposition and an adverb. Then a special rule called "Preposition-Adverb" is launched, which makes the final choice. Since this rule is triggered after the formation of the nominal group, the case check is performed at the centre of the nominal group, which ensures the correct choice of these two tokens.

Table 3 shows the most common prepositional phrases of the second group, which require the genitive case after them.

To calculate the frequency of formation of each preposition, here and further, about 300 sentences from the main body of the NCRL had been used, supplemented, if necessary, by the sentences of the newspaper body and the available array of texts (of the volume of about 50 million words), composed of a number of stories, news and sports articles. The selected material had undergone an additional filtering to exclude cases of punctuation marks breaking the phrase (in this case it is definitely not a prepositional turn). Then the automatic analysis of the selected sentences was launched. The obtained result was saved as an xml-file that was finally used to determine the frequency of occurrence of specific preposition.

Table 3.
The most frequent phrases of the second group

Turn of speech	Link with the main word	Frequency, ipm	From which preposition
В КОНЦЕ (‘at the end of’)	Где, Когда (‘Where’, ‘When’)	162.2	92%
В НАЧАЛЕ (‘at the beginning of’)	Где, Когда (‘Where’, ‘When’)	83.0	93%
В ПОДТВЕРЖДЕНИЕ (‘in confirmation of’)	Как (‘How’)	3.5	76%
В РАМКАХ (‘within’)	Как (‘How’)	31.6	95%
В СЕРЕДИНЕ (‘in the middle of’)	Где, Когда (‘Where’, ‘When’)	29.4	89%
ВО ГЛАВЕ (‘headed by’)	Где (‘Where’)	40.4	69%
К КОНЦУ (‘towards the end’)	Когда, Куда (‘When’, ‘Where’)	36.0	78%
К НАЧАЛУ (‘towards the beginning of’)	Когда, Куда (‘When’, ‘Where’)	11.2	93%
НА КРАЮ (‘on the verge’)	Где (‘Where’)	13.6	87%
НЕ СЧИТАЯ (‘not counting’)	Как (‘How’)	7.7	84%
ПО АДРЕСУ (‘about’)	Как (‘How’)	10.6	18%
ПО ПОРУЧЕНИЮ (‘on the instructions of’)	Почему (‘Why’)	4.8	96%
ПО ПРАВУ (‘by right’)	Почему (‘Why’)	7.1	17,5%
ПО ПРОСЬБЕ (‘at smb’s request’)	Почему (‘Why’)	7.6	94%
ПО СЛУЧАЮ (‘on the occasion of’)	Почему (‘Why’)	25.1	79%
СО СТОРОНЫ (‘on smb’s part’)	Откуда (‘From where’)	90.7	72%0

4. Group 3. Collocations that may not be phrasemes

This group includes complex homonymous prepositional phrases, whose phrasemes can serve as prepositions or be a simple combination of words. In the first case, all the words that form the phrase must be combined into a single token, in the second case, they must be left unchanged. Thus, the pre-syntactic module, having marked out the next phrase belonging to the third group, cannot combine its tokens into a single one by itself. For further processing of the phraseme, a rule that is practically the first in succession is launched, deciding whether this phraseme may be a prepositional phrase or not. In our dictionary there are about 90 phrasemes of this kind. It should be noted that at the stage of parser analysis, the nominal groups are not yet formed, that is why the rules for analysing these phrasemes are significantly complicated.

The most detailed description of such collocations is given in a study by Rogozhnikova [4], who analyzes them from a semantic point of view. However, this semantics is considered from the point of view of a "person", not a "computer", so it lacks strict formal features. Therefore, when developing rules for text processing, we have to take into account only the surrounding context, its grammar and classes. Sometimes we also have to take into consideration the remote context.

In connection with this approach, it is possible to divide the prepositional turns of this group into 3 subgroups, depending on the complexity of their analysis.

4.1. Subgroup 3A

This subgroup includes homonymous phrasemes, which can play a role of a preposition if the simplest criterion is fulfilled. This criterion is the presence of a word on the right in the genitive case. In the event of absence of such a case, the phraseme remains a simple combination of words. Example:

● *Онтологические системы могут использоваться для решения различных* ***задач в сфере*** *искусственного интеллекта* ('Ontological systems can be used to solve various problems in the field of artificial intelligence').

● *В сфере радиусом в 100 световых лет насчитывается около 10000 звёзд* ('There are about 10,000 stars in a sphere with a radius of 100 light years').

This subgroup includes 13 turns, the most common ones are presented in Table 4. As before, there are prepositional phrases that can be connected to the main word with various links. Example:

●*Такие счета могут быть номинированы в иностранной валюте, а владельцы счёта NRI могут определять бенефициария* ***в пределах*** *Индии* ('Such accounts can be denominated in a foreign currency, and NRI account holders can assign a beneficiary within India.') – (определить – Где (Директив)– *в пределах*) (Where (Directive)).

●*Отступления сделаны для пироксенов, гранатов, хлоритов и амфиболов, поскольку минералы* ***в пределах*** *этих групп близки по условиям формирования...* ('Deviations are made for pyroxenes, garnets, chlorites, and amphiboles, since the minerals within these groups are similar in terms of formation conditions...') – (*близки* – Как (Характеристика способа или меры действия) – *в пределах*) (How (Description of method or measure of an action)).

Table 4.
The most frequent phrases of subgroup 3A

Turn of speech	Link with the main word	Frequency, ipm	From which preposition
В ПРЕДЕЛАХ ('within the limits of')	Где, Как ('Where', 'How')	23.8	76%
В СЛУЧАЕ ('in the event of')	Когда ('When')	71.4	67%
В СФЕРЕ ('in the field of')	Как ('How')	15.4	96%
В ЧИСЛЕ ('in the number of')	вПред ('Prepositional')	32.3	88%
В ЧИСЛО ('to the number of')	вВин ('Accusative')	7.6	81%
С ЦЕЛЬЮ ('with a view to')	Зачем ('For what reason')	30.2	65%

4.2. Subgroup 3B

This subgroup includes homonymous phrasemes, which can play a role of a preposition or remain a simple combination of words. To select a particular option, it is necessary to fulfil a complex condition. To have it implemented, the surrounding context, grammar, and classes of individual words have to be taken into account. Sometimes it is necessary to take into consideration even the remote context within the entire sentence [16].

For example, the phrase *на глазах у* ('before smb's eyes') can have two semantic meanings: something happens to somebody's eyes (and this will be a free combination of three words) or something happens in the presence of someone (and this will be a prepositional phrase). To analyse such a phraseme, the following rule is used: if one of the following words – *влага* ('moisture'), *слеза*

(‘tear’), *слезы* (‘tears’) – occurs to the left or right of the phraseme within seven words from it, then we deal with a simple word combination, otherwise it is a prepositional phrase. It has to be noted that in both cases, the phrase is followed by a word in the genitive case:

● *На глазах у Маруси появились слезы* (‘Marusia's eyes filled with tears’).

● ***На глазах у*** *посетителей, так и не слезших со столов, ему удалось поймать 28 змей* (‘Before the very eyes of the customers, who had not got off the tables, he managed to catch 28 snakes’).

This subgroup includes about 60 turns, the most common ones are presented in Table 5. As before, there are some prepositional phrases that can be connected to the main word by various links. For example, for the phrase *по вопросу* (‘on the issue of’):

● *Заседание Госдумы* ***по вопросу*** *его ратификации состоится 20 или 21 марта* (‘The State Duma will hold a meeting on its ratification on March 20 or 21’) (*Заседание* – Какой – *по вопросу*) (Which).

● *Самым ярым оппонентом Кука* ***по вопросу*** *распространения американских культурных растений в области Тихого океана много лет был его соотечественник Меррилл* (‘Cook's most ardent opponent on the issue of the distribution of American cultivated plants in the Pacific region for many years was his compatriot Merrill’) (*оппонентом* – поДат – *по вопросу*) (Dat).

По вопросу губернатора Резанов догадался, что тот значительно больше его осведомлен (‘The governor's question made Riazanov guess that the latter was much more knowledgeable than he was’) – a simple combination of a preposition and a noun.

Table 5.
The most frequent phrases of subgroup 3B

Turn of speech	Link with the main word	Frequency, ipm	From which preposition
В ГЛАЗАХ (‘in smb’s eyes’)	вПред (‘Prepositional’)	50.5	44%
В КАЧЕСТВЕ (‘as’)	Как (‘How’)	99.8	94%
В ОБЛАСТИ (‘in the field of’)	вПред (‘Prepositional’)	53.5	79%
В ОТНОШЕНИИ (‘with respect to’)	вПред (‘Prepositional’)	48.7	84%
В ПОРЯДКЕ (‘by way of’)	Как (‘How’)	42.1	29%
В РАЙОНЕ (‘around’)	Где (‘Where’)	34.0	10%
В СИЛУ (‘because of’)	Как (‘How’)	41.7	74%
С ПОМОЩЬЮ (‘with the help of’)	Как (‘How’)	68.5	57%
С ТОЧКИ ЗРЕНИЯ (‘from the point of view of’)	Как (‘How’)	33.7	98%

4.3. Subgroup 3C

This subgroup includes the most complex homonymous phrasemes, which can play a role of a preposition, an adverb, or remain a simple combination of words. To select a particular option, a rather lengthy criterion has to be fulfilled. In general, to have it implemented, the surrounding context, grammar, and classes of individual words have to be taken into account. Sometimes it is necessary to take into consideration even the remote context within the entire sentence. For example, let us examine the phraseme *в результате* (‘as a result’). In Rogozhnikova's study, some semantic justification and examples are provided [4]. With this basis, the following rule has been developed for the analysis of the phraseme.

If there are the lemmas СОМНЕВАТЬСЯ (‘to doubt’), СОМНЕНИЕ (‘doubt’), УВЕРЕННЫЙ (‘confident’) to the left of the phraseme and if there are lemmas АНАЛИЗ (‘analysis’), ГОЛОСОВАНИЕ (‘voting’), ИССЛЕДОВАНИЕ (‘research’), ОПЕРАЦИЯ (‘operation’), ОПЫТ

(‘experience’), ТЕСТ (‘test’), ЭКСПЕРИМЕНТ (‘experiment’) to the right (directly or in one word in the genitive case), the wordforms of which are in the genitive case, then the phraseme is a simple combination of words. Example:

● *Не будучи уверен в результате голосования и не желая идти на риск и в то же время сильно надеясь на воздействие ленинской речи, левый блок сделал уступку…* (‘Not being sure of the vote results and unwilling to take any risks, and at the same time pinning great hopes on the impact of Lenin's speech, the left bloc made a concession…’)

● *Я не сомневался в результате этого эксперимента* (‘I had no doubts about the result of this experiment’).

If there are the lemmas СОМНЕВАТЬСЯ (‘to doubt’), СОМНЕНИЕ (‘doubt’), УВЕРЕННЫЙ (‘confident’) to the left of the phraseme and a comma or a full stop to the right of it, then the phraseme is also a simple combination of words:

● *Мой добрый друг был, как правило, уверен в результате* (‘My good friend was generally confident of the result’).

If there is a word in the genitive case to the right of the phraseme, then it performs the function of a preposition:

● *Это сообщение выдаётся автоматизированной системой, если* ***в результате*** *вычисления формула получила значение "ложь"* (‘This message is issued by the automated system if the formula has received the value "false" as a result of the calculation’).

Otherwise, the phraseme performs the function of an adverb:

● ***В результате*** *объекты имитационной модели перейдут в некорректные состояния* (‘As a result, the objects of the simulation model will come to incorrect states’).

Thus, it is clear that there is a possibility to formalize semantic relations, but sometimes this process results in rather lengthy rules.

Today, this subgroup includes 15 phrases, the most common ones are shown in Table 6. It should be noted that at least two of them have more than three variants of homonymy. Thereby, the phraseme *в меру* (‘within reasonable limits’) can additionally be a predicate: *Вроде бы все* ***в меру****, все на своих местах* (‘Everything seems to be within reasonable limits, everything is in its place’). The phraseme *в разрезе* (‘in section’) can additionally perform the function of an attribute: *У меня над кроватью, сколько себя помню, висел план огромного океанского парохода* ***в разрезе*** (‘I have had a plan of a huge ocean steamship in section hanging over my bed for as long as I can remember’).

Table 6.
The most frequent phrases of subgroup 3C

Turn of speech	Link with the main word	Frequency, ipm	From which preposition
В ЗАКЛЮЧЕНИЕ (‘in conclusion’)	Где, Когда (‘Where’, ‘When’)	14.7	34%
В МЕРУ (‘within reasonable limits’)	Как (‘How’)	9.5	46%
В РЕЗУЛЬТАТЕ (‘as a result of’)	Как (‘How’)	81.2	62%
ЗА РАМКИ (‘exceeding the limits of’)	Куда (‘Where’)	3.8	95%
НА РАССТОЯНИИ (‘away from’)	Где (‘Where’)	12.8	42%
НА СТОРОНЕ (‘on smb’s side’)	Где (‘Where’)	11.5	70%
НА ФОНЕ (‘against a background’)	Как (‘How’)	24.1	75%
ПО ОКОНЧАНИИ (‘after’)	Когда (‘When’)	19.5	93%
ПО ПУТИ (‘on the way’)	Куда (‘Where’)	18.8	35%

5. Conclusion

As a result of the study, the classification of turns of speech (phrasemes) has been performed depending on the type of homonymy. Rules have been developed that allow to remove part-of-speech and syntactic homonymy with high accuracy. We believe that due to the large variability of the Russian language, raising the accuracy of parsing a certain number of constructions to the level of above 95% may require disproportionately large efforts, and, in fact, may turn into to analysing specific phrases. Therefore, in some cases, rarely encountered phrasemes have been ignored. For example, the construction *под знаком* + *род. пад* (‘under the sign of’ + genitive case) may occur in the main and newspaper corpora of the NCRL over 1700 times, while only 9 cases turned out to be free word combinations, and not compound prepositions (*под знаком интеграла*… (‘under the sign of the integral...’)).

At the same time, the removal of semantic homonymy is a much more complex task that requires additional research.

6. References

[1] M.V. Kopotev, T.I. Steksova, Isklyuchenie kak pravilo: Perekhodnye edinicy v grammatike i slovare. M.: Yazyki slavyanskoj kul'tury: Rukopisnye pamyatniki Drevnej Rusi, 2016. (In Russian).
[2] National Corpus of the Russian Language. URL: http://www.ruscorpora.ru/. (In Russian).
[3] S.A. Kuznetsov Bol'shoy tolkoviy slovar russkogo yazika. SPb.: Norint, 1998. (In Russian).
[4] R.P. Rogozhnikova Tolkovyj slovar' sochetanij, ekvivalentnyh slovu. M.: OOO «Izdatel'stvo Astrel'», 2003. (In Russian).
[5] V. Zakharov, A. Golovina, E. Alexeeva, V. Gudkov Russian Secondary Prepositions: Methodology of Analysis, XVI Mezhdunarodnaya konferenciya po komp'yuternoj i kognitivnoj lingvistike (TEL 2020).
[6] L. Iomdin , V. Petrochenkov, V. Sizov, L. Tsinman, Etap parser: state of the art. Computational Linguistics and Intellectual Technologies. Based on the materials of the annual international conference "Dialogue" (Bekasovo, May 30 - June 3, 2012), issue 11 (18), Moscow: RGGU Publishing House, 2012. vol. 2, pp. 117–131.
[7] K.V. Anisimovich, K.Ju. Druzhkin, F.R. Minlos, M.A. Petrova, V.P. Selegey, K.A. Zuev, Syntactic and semantic parser based on ABBYY Compreno linguistic technologies // Computational Linguistics and Intellectual Technologies. Based on the materials of the annual international conference "Dialogue" (Bekasovo, May 30 - June 3, 2012), issue 11 (18), Moscow: RGGU Publishing House, 2012. vol. 2, pp. 91–103.
[8] Linguistic processor ETAP-4, URL: http://www. http://proling.iitp.ru/ru/etap4.
[9] K.K. Boyarsky, E.A. Kanevsky, Semantiko-sintaksicheskiy parser SEMSIN, Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2015, vol. 15, №5, pp. 869–876. (In Russian).
[10] K.K. Boyarsky, E.A. Kanevsky, Slovosochetaniya, ekvivalentnye slovu, International Conference "Internet and Modern Society" (IMS-2015) – SPb, ITMO University, 2015, pp. 55–66. (In Russian).
[11] V.A. Tuzov, Komp'yuternaya semantika russkogo yazyka. SPb: SPbU. Publishing House, 2004. (In Russian).
[12] K.K. Boyarsky, E.A. Kanevsky, S.K. Stafeev, Ispol'zovanie slovarnoj informacii pri analize teksta, Scientific and Technical Journal of Information Technologies, Mechanics and Optics. 2012, №3 (79), pp. 87–91. (In Russian).
[13] G.A. Zolotova, Sintaksicheskij slovar'. Moscow: Editorial URSS, 2011. (In Russian).
[14] V. Zakharov, K. Boyarsky, A. Golovina, A. Kozlova, Semantic Analysis of Russian Prepositional Constructions, RASLAN 2020. Recent Advances in Slavonic Natural Language Processing. Proceedings. Brno, 2020, pp. 103–113.
[15] E.A. Kanevskij, E.N. Klimenko, E.F. Silina, Osobye narechnye oboroty, Vtorye chteniya pamyati professora B.L. Ovsievicha «Ekonomiko-matematicheskie issledovaniya:

matematicheskie modeli i informacionnye tekhnologii»: Materialy Vserossijskoj konferencii. – SPb.: Nestor-Istoriya, 2015, pp. 101–107. (In Russian).

[16] E.A. Kanevsky, Osobye predlozhnye oboroty, Kontrastivnye issledovaniya i prikladnaya lingvistika: mater. Internat. sci. conf., Minsk, 2014, part 1. Minsk: MGLU, 2015, pp. 115–119. (In Russian).