

# Comparative Corpus-driven Study of Prepositional Semantics in Russian and Czech

Victor Zakharov

Saint-Petersburg University

**Abstract.** This paper deals with prepositions with causal meaning in Russian and Czech. In Slavonic languages prepositions are closely connected to cases. Russian and Czech prepositions have many common features. Prepositions show a relation in space or time or a special relationship between two or more people, places, things or situation. In the current paper we are dealing with causal relations. There are different ways to express them. Among these means, the most common are prepositional-case forms and complex sentences with a subordinate causal part. We analyze the repertoire of causal prepositions in both languages and describe their statistical representation in corpora. Another task is to reveal translation equivalents between two languages.

**Keywords:** preposition, causal meaning, Russian language, Czech language, corpus statistics, parallel corpora

## 1 INTRODUCTION

The preposition is perhaps the most mysterious part of speech in all languages. Its frequency is extraordinarily high. In Russian and Czech, more than 10% of tokens in a given text are prepositions (Lyashevskaya and Sharov, 2009; Statistika čestiny, 2009). Prepositions are heterogeneous in both languages: there is a small group of primary prepositions (about 30) and a few hundreds of secondary ones. The latter are motivated by content words (nouns, adverbs, verbs), which may be combined with primary prepositions forming complex multiword expressions. But a strict division between secondary multiword prepositions and prepositional phrases is not specified. This is a task for a special corpus-based research.

The semantics of prepositions is a special topic of interest. Primary prepositions tend to be highly polysemous. For instance, the preposition "в" in Russian ('in') has 23 meanings in the explanatory dictionary (Dictionary..., 1988). The Czech preposition "v" ('in') has 16 meanings (Slovník..., 1977). The majority of these meanings are, however, quite rare. There are also many unique or "empty" meanings when the preposition is part of an idiom. Prepositional ambiguity is manifested in the complex nature of the prepositional meaning and in selective preferences of certain prepositions, depending on context.

Prepositions are characterized as function words used to express various relationships between main (governor) and dependent (governee) members of a phrase. The difficulty is that the relations expressed by prepositions are multi-sided, grammatical, lexical and extralinguistic. Prepositions are often said to have no real lexical meaning. At the same time, they express different semantic relations between words, and their meanings must naturally directly correspond to these relations.

We believe that the meaning of prepositions is realized in prepositional constructions. In order to describe prepositional meaning, it is necessary to describe the meaning of a prepositional construction. It is important to provide a special metalanguage for such a description. For Russian, we use the syntaxeme classification of G. Zolotova (Zolotova, 2011). The syntaxeme combines the governor, preposition and governee in a case form and is a minimal semantic-grammatical unit. The designation of syntaxemes in (Zolotova, 2011) is formed according to the pattern of semantic roles: directive, destinative, correlative, quantitative, qualitative, locative, mediative, temporative; 27 syntaxemes in total (Zolotova, 2011, p. 383). All grammars of Czech describe prepositional meanings autonomously in terms of adverbial modifiers. As a rule, 7 types of such meanings are distinguished (Štícha, 2018).

## 2 PREPOSITIONS WITH CAUSAL MEANING IN RUSSIAN AND CZECH

For this study we chose Czech and Russian prepositions with similar meanings, i.e., causal relationship (příčinný vztah in Czech). This is the meaning of constructions where the prepositional group indicates the cause of an action or the influencing factor. The word 'cause' is the basic term used to interpret the whole lexical composition, which is associated with the category of determining the cause. Dictionaries of a language divide two meanings of this word: 1) cause as a phenomenon that inadvertently causes another phenomenon, ontological cause; 2) cause as a basis, precondition for the realization of an event, action, i.e., subjective, explainable cause. The meaning of prepositions is based on that description. Among ways to express causal relations, the most common are prepositional-case forms and complex sentences with a subordinate causal part, and most of causal conjunctions come from prepositions.

A lot of studies are devoted to this type of relation (Křížková, 1967; Vsevolodova, Yascenko, 1988; Horák, 1989; Diessel, Hetterle 2011) and only several studies deal with causal prepositions (Kroupová, 1980; Levontina, 1997; Iordanskayaca, Melchuk, 1996; Luraghi 2005). However, these studies were not based on corpora. Our study relies on statistics on the use of these prepositions in large text material.

Causal relationships can be expressed in Russian and Czech by some primary prepositions and a large number of secondary ones. They can enter into connection with nouns and pronouns in genitive, dative, accusative and instrumental. The lists of causal prepositions differ in different sources. For Czech we started from (Štícha, 2018) with a few deviations. The main part of this list is as follows: *z, za, pro, skrz, od, na, v důsledku* 'due to', *následkem* 'due to', *za příčinou* 'for the reason of', *vinou* 'due to', *díky* 'thanks to', *kvůli* 'due to', *z důvodu* 'for reasons of', *u příležitosti* 'on the occasion of', *v souvislosti s* 'in relation to', *v závislosti na* 'depending on', *vlivem* 'owing to', *vycházejí/jí z* 'coming from'. In Russian, according to our analysis, the list is as follows: *za, u3, u3-*

*за, на, от, по, под, после, при, с, через, благодаря* ‘thanks to’, *в зависимости от* ‘depending on’, *в ответ на* ‘in response to’, *в результате* ‘as a result of’, *в свете* ‘in light of’, *в связи с* ‘due to’, *в силу* ‘by force of’, *за счёт* ‘on account of’, *исходя из* ‘drawing from’, *на основании* ‘on the basis of’, *на основе* ‘based on’, *на почве* ‘on the ground of’, *по причине* ‘because of, for the reason of’. (Here we provide English equivalents only for the secondary prepositions because the meanings of the primary ones are highly context-dependent).

These prepositions form clusters of intralanguage and interlanguage synonymy. Different prepositions can express the same meanings and grammatical relations when used in the same phrases. In the sentences *Он не пришел по причине болезни – Он не пришел из-за болезни – Он не пришел вследствие болезни* (‘He did not come because of the disease - He did not come due to the disease - He did not come on account of the disease’) it is possible to interpret the prepositions as synonyms. In Czech, equivalent synonymous groups are formed by groups ‘*v důsledku/kvůli/pro*’ or ‘*kvůli/v důsledku/následkem/ z důvodu*’.

Consequently, our tasks in the current study were as follows:

- 1) 1) obtain statistical characteristics of prepositions from corpora;
- 2) show differences in use of prepositions in various genres;
- 3) reveal main translation equivalents for the basic set of prepositions with the causal meaning.

### 3 RESULTS AND DISCUSSION

The results of the statistical analysis presented in this article were acquired on the Araneum Russicum III Maius and Araneum Bohemicum IV corpora (1.25 billion tokens each, [www.unesco.uniba.sk](http://www.unesco.uniba.sk)), the Russian National Corpus (RNC) (321 milion tokens, [www.ruscorpora.ru](http://www.ruscorpora.ru)), the Czech National Corpus (ČNK, [www.korpus.cz](http://www.korpus.cz)) - syn v8 corpus, 5,4 bilion tokens, and InterCorp v13. InterCorp includes parallel corpora for 40 languages. We used the Russian-Czech and Czech-Russian parallel subcorpus which consists of both Russian texts translated into Czech and Czech texts translated into Russian. Moreover, it also includes texts translated into Czech and Russian from other languages, most often from English (Rosen et al., 2020).

Prepositional causal constructions occupy a rather modest place among all prepositional constructions. In a large experiment on the extraction and annotation of Russian prepositional constructions, only 349 causal constructions were identified among 10047 constructions, i.e., 3.47%. Their analysis showed that in Russian, most prepositional constructions with the causal meaning are represented by constructions with the following prepositions: *по* (24% of all occurrences inside causal constructions), *за* (19%), *из-за* (16%), *от* (12%), *в связи с* (6%), *на*

(5%), на основании/на основе (4%), по причине (3%). For Czech, such an extensive experiment is yet to be undertaken, and so more or less accurate statistics cannot be provided currently.

However, for most of the **primary** polysemantic prepositions that form the majority of causal constructions this meaning is not the main one. In order to identify the prepositions for which the causal meaning is the only or the main one, we conducted another experiment. As a rule, these are secondary multiword prepositions. 12 Russian and 11 Czech prepositions mostly from the top part of the frequency list were selected and constructions with them were extracted from two Russian and two Czech corpora. Then the first 50 randomly selected constructions were annotated by hand according to the realized meaning. Usually, in those cases where this meaning differed from the causal one we were dealing with a free combination of a preposition and a common word. The assignment of these combinations to the category of prepositions is sometimes complicated by the fact that the common word has to some extent retained its original lexical meaning.

The results obtained are shown in Tables 1 and 2. They demonstrate relative frequencies of each preposition and the percent of constructions with causal meaning.

| Preposition           | Frequency (ipm) in Araneum | % of prepositional use, Araneum | Frequency (ipm) in RNC | % of prepositional use, RNC |
|-----------------------|----------------------------|---------------------------------|------------------------|-----------------------------|
| в зависимости от(ото) | 111,60                     | 100                             | 27,72                  | 98                          |
| исходя из(изо)        | 44,70                      | 100                             | 14,95                  | 100                         |
| за счёт (за счѐт)     | 142,57                     | 88                              | 41,70                  | 100                         |
| в связи с(со)         | 119,10                     | 100                             | 57,93                  | 100                         |
| на основе             | 105,40                     | 82                              | 35,46                  | 82                          |
| на основании          | 57,30                      | 100                             | 31,78                  | 100                         |
| в результате          | 173,80                     | 100                             | 81,26                  | 54                          |
| по причине            | 19,70                      | 86                              | 12,34                  | 96                          |
| в ответ на            | 12,20                      | 100                             | 18,09                  | 98                          |
| на почве              | 1,10                       | 100                             | 3,59                   | 96                          |
| в свете               | 2,00                       | 84                              | 3,68                   | 78                          |
| в силу                | 10,70                      | 90                              | 10,54                  | 80                          |

**Table 1.** Relative frequency (ipm) and percentage of prepositional use of Russian multiword prepositions in the Araneum Russicum III Maius and in the RNC.

| Preposition     | Frequency (ipm) in Araneum | % of prepositional use, Araneum | Frequency (ipm) in ČNK | % of prepositional use, ČNK |
|-----------------|----------------------------|---------------------------------|------------------------|-----------------------------|
| díky            | 313,30                     | 100                             | 186,20                 | 100                         |
| kvůli           | 225,60                     | 100                             | 215,40                 | 100                         |
| následkem       | 5,50                       | 100                             | 5,70                   | 100                         |
| v důsledku      | 26,50                      | 100                             | 26,60                  | 100                         |
| v souvislosti   | 41,30                      | 92                              | 28,80                  | 92                          |
| v závislosti na | 23,00                      | 100                             | 16,80                  | 100                         |
| vinou           | 3,10                       | 96                              | 8,40                   | 98                          |
| vlivem          | 17,00                      | 66                              | 17,40                  | 80                          |
| z důvodu        | 43,80                      | 96                              | 15,70                  | 100                         |

|                |              |     |      |     |
|----------------|--------------|-----|------|-----|
| za příčinou    | 0,10         | 100 | 0,10 | 100 |
| vycházejí/jí z | only 3 items | 66  | -    | -   |

**Table 2.** Relative frequency (ipm) and percentage of prepositional use of Czech secondary prepositions in the Araneum Bohemicum IV Maius and in the ČNK.

Different corpora for both languages, one a well-balanced national corpus and the second a web-based corpus, were selected on purpose. We wanted to make sure that the choice of corpora has no significant effect on the final results. The result is ambiguous. We see that national corpora show generally less frequency than web-corpora. This is quite clear, since their percentage of journalistic and business texts for which secondary prepositions are more characteristic is smaller.

Causal prepositions, especially secondary ones, occur more frequently in journalistic and theoretical writing. We have observed the distribution of causal prepositions in texts of various genres and types. Our data and data from Czech scientific publications indicate that the relative frequency of causal prepositions is highest in the texts on legislation and journalism. From there, however, they also penetrate colloquial language or fiction.

#### 4 TRANSLATION EQUIVALENTS

In most cases, Russian primary causal prepositions do not correspond to their similar Czech counterparts. On the one hand, the repertoire of prepositions is different in both languages, and on the other hand, they are used differently. These are either differences in the semantic structure of the prepositions or differences in individual use cases.

Actually, only Russian *из* 'out of' and *за* 'for'<sup>1</sup> coincide with the Czech *z* and *za*: *сделать что-нибудь из расчета – udělat něco z vypočítavosti* ('to do something out of self-seeking'); *наказывать за проступок – trestat za provinění* ('to be punished for a wrongdoing'). Such coincidence, however, does not always take place, e.g.: *сердиться за упрёк – zlobit se pro výtku* ('to be angry for a reproach').

The preposition *от* 'from' is highly polysemous in Russian and occurs very often in various meanings. Therefore, translation of constructions with this preposition may differ: *от болезни – na nemoc* 'from illness', *от наводнения – následkem povodni* 'due to flood', *от этого слова – z tohoto slova* 'from this word', *произошло не от хорошей жизни – nedošlo v důsledku dobrého života* 'did not come from a good life'.

In Russian, this preposition is used in conjunction with the name of the disease: *умер от рака, от сердечного приступа*. In such cases, the preposition *na* with the accusative is used in Czech: *zemřel na rakovinu, na infarkt* 'he died of cancer, of a heart attack'. It could be translated as

<sup>1</sup> Only a rough translation for prepositions mentioned in the paper henceforth can be provided out of context.

*v důsledku* or *následkem* ‘as a result, as a consequence’ when the cause is not directly related to a consequence that has already occurred in the meantime.

The Czech preposition depends often on **the type of the noun** entering the prepositional phrase and sometimes on the verb. If it is a noun with the meaning of inner feeling, state of spirit, the Russian expression corresponds not to the prepositional construction in Czech but the corresponding noun in the instrumental: *плакать om радости* - *plakat radosti* ‘weep for joy’. However, constructions with nouns of other semantics are translated sometimes in a similar way, too (cf. : *заболеть om шума* – *onemocnět hlukem* ‘to get sick from noise’, *погибнуть om рук* – *zahynout rukou* ‘to die at the hands of’).

The choice of the Czech equivalent can also depend on the meaning of the governing verb. The construction *зависеть om чего-нибудь* ‘depend on something’ is used always with the genitive of the noun. The governee is then a proverbial determination of the cause. In Czech, the verb *záviset* ‘depend’ is associated with the preposition *na* ‘on’ and the noun in the locative: *зависеть om укрепления демократии* - *záviset na upevnění demokracie* ‘depend on strengthening democracy’. The rules of both languages therefore do not allow for other variants of translation of this construction.

However, sometimes different translations of the same prepositional construction are possible. They are determined by the style, preference of a translator and the prevailing practice. We attempted to identify some Russian-Czech prepositional equivalents in causal constructions. Sometimes these correspondences are quite frequent, sometimes not.

As the first step of the study we analyse the translation of a few Russian preposition on the base of the InterCorp Russian-Czech corpus (Rajnochová et al., 2020).

The frequency of prepositions in InterCorp is high, and in the absence of semantic annotation, the task of selecting causal prepositional constructions is nontrivial. For this, the main lexical markers (nouns) were manually identified in constructions with the analyzed prepositions in a causal meaning. Then these words were included in the CQL query, taking into account the required dependent noun case. Then manual cleaning of the list of extracted constructions was performed.

It should also be emphasized that InterCorp is mainly composed of literary texts and film transcripts. Translators did not always translate Russian prepositional constructions into Czech ones, translation is often free.

The following preliminary results of analysing translation equivalents were obtained. As mentioned previously, the preposition *уз* (*uzo*) is one of the few that has a direct Czech equivalent, it is the preposition *z* (*ze*). From the selected 212 *uz*-constructions with a total frequency of 1127,

some constructions with a frequency greater than 12 were selected. The results are shown in Table 3.

| Construction                        | Frequency in corpus | The number of causal meanings | Translation using the preposition z(ze) | Another causal preposition <sup>2</sup> or periphrasis <sup>3</sup> | Translation by other means <sup>4</sup> |
|-------------------------------------|---------------------|-------------------------------|-----------------------------------------|---------------------------------------------------------------------|-----------------------------------------|
| из вежливости 'out of courtesy'     | 40                  | 39                            | 24                                      | 2                                                                   | 15                                      |
| из безопасности 'out of safety'     | 37                  | 15                            | 12                                      | 1                                                                   | 2                                       |
| из жалости 'out of pity'            | 24                  | 24                            | 14                                      | 6                                                                   | 4                                       |
| из ненависти 'out of hatred'        | 17                  | 16                            | 9                                       | 5                                                                   | 2                                       |
| из необходимости 'out of necessity' | 16                  | 10                            | 7                                       | 1                                                                   | 2                                       |
| из мести 'out of revenge'           | 13                  | 13                            | 7                                       | 3                                                                   | 3                                       |
| из гордости 'out of pride'          | 13                  | 12                            | 6                                       | 6                                                                   | 0                                       |

**Table 3.** Translation equivalents for Russian preposition *из* (*изо*)

Thus, the analysis showed that indeed the preposition *из* in causal meaning is mainly translated by the preposition *z* and only occasionally by other prepositions.

A noticeably different picture is as concerns the preposition *из-за*. In the 280 analyzed causal constructions, it is translated by a variety of prepositions: *kvůli*, *díky*, *pro*, *z* (*ze*), *vzhledem*, *výsledkem*, *důsledkem* (in descending order of the frequency of the equivalents).

Then secondary multiword prepositions including the primary preposition *в* 'in' were analyzed. The preliminary analysis results are shown in Table 4. Numbers in cells show the number of translation equivalents for the corresponding Russian preposition in the column.

| Czech preposition   | в зависи-<br>мости от(о)    | в ответ на | в<br>результате | в связи<br>с(о) | в силу |
|---------------------|-----------------------------|------------|-----------------|-----------------|--------|
|                     | the number of constructions |            |                 |                 |        |
|                     | 260                         | 201        | 947             | 489             | 145    |
|                     | The number of translations  |            |                 |                 |        |
| v závislosti na     | 11                          |            |                 |                 |        |
| podle (dle)         | 123 (5)                     |            |                 |                 |        |
| v (jako) odpověď na |                             | 14 (3)     |                 |                 |        |
| výsledkem           |                             |            | 13              |                 |        |
| následkem           |                             |            | 10              |                 |        |
| v důsledku (-em)    |                             |            | 162 (18)        |                 | 1 (3)  |

2 из предосторожности - *kvůli* vší opatrnosti 'for all caution'

3 хранитель печати взглянул на меня **из вежливости** 'the seal superior looked at me out of **politeness**' - představený pečeti se ohlédl **zdvořile** po mně 'the seal superior looked at me **politely**'

4 из жалости и сострадания - **máme soucit**, slitovali jsme se máme soucit, slitovali jsme se 'we have compassion, we have pity'

|                            |  |  |    |       |    |
|----------------------------|--|--|----|-------|----|
| díky                       |  |  | 31 |       | 14 |
| po                         |  |  | 35 |       |    |
| při                        |  |  | 34 | 7     |    |
| pro                        |  |  |    | 5     |    |
| v souvislosti s            |  |  |    | 63    |    |
| vzhledem k                 |  |  |    | 17    | 4  |
| ve vztahu                  |  |  |    | 5     |    |
| ve spojitosti (ve spojení) |  |  |    | 8 (2) |    |
| kvůli                      |  |  |    | 57    |    |
| z                          |  |  |    |       | 19 |
| podle                      |  |  |    |       | 7  |

**Table 4.** Translation equivalents for Russian multiword prepositions with causal meaning

The table does not include prepositional equivalents with frequencies of less than 5: *vinou, u příležitosti, na základě*. Let's note that the preposition *в зависимости от* 'depending on' was often translated by the construction *záleží na*, precisely for 14 times.

We see that sometimes there are frequent variants of translation of certain prepositions. The presence of several translations makes it possible, apparently, to speak of their synonymy. It should be also emphasized that Russian prepositional constructions were often translated into Czech by subordinate clauses or other phrases. However, this is the task for a separate study.

In a number of synonymous series, the asymmetry of translation equivalents should be noted when translating from Russian into Czech and from Czech into Russian. It is not clear whether this applies to linguistics, or more to translatology. A more detailed and in-depth analysis is yet to come.

## 5 CONCLUSION

To conclude, it can be noted that causal meanings in prepositional constructions in both Russian and Czech are expressed by a few primary prepositions and a large number of secondary ones. The use of secondary multiword prepositions to express the determination of the cause is very rich in Russian, as evidenced by their frequent and diverse representation in journalistic-style texts. In Czech, the number of secondary prepositions is not as high. In addition, they are often synonymous with each other and do not have the same semantic variety as Russian prepositions. For this reason, it is often difficult to find an equal counterpart to the Russian prepositional construction.

We would like to emphasize that it is important to set and gradually solve the task of comparing not just individual prepositions in two languages, but the prepositional systems as a whole. Such an analysis makes it possible to define study of prepositional meanings as contribution to linguistic theory as well as useful for practical tasks.

## ACKNOWLEDGEMENTS



This work was supported by the Russian Foundation for Basic Research (grant No. 17-29-09159 “Quantitative grammar of Russian prepositional constructions”).

We express our sincere gratitude to the 2<sup>nd</sup> year students of the SPbU Mathematical Linguistics Department for their valuable help in annotating the data.

Our cordial thanks belong to Vladimir Benko and Alexandr Rosen for their help in work with corpora.

We also thank Anastasia Golovina for her help in checking the English text of the paper.

## REFERENCES

- DIESEL, Holger —HETTERLE, Katja: Causal clauses: A crosslinguistic investigation of their structure, meaning, and use. In: P. Siemund (ed.). *Linguistic Universals and Language Variation*. Berlin, Boston: De Gruyter Mouton, 2011, p. 21 – 52.
- HAVRÁNEK, Bohuslav (ed. a kol.): *Slovník spisovného jazyka českého*. Praha: Academia 1960-1971.
- HORÁK, E.: Východiská pre konfrontáciu predložkového systému slovenčiny s inými jazykmi. In: *Studia Academica Slovaca*. 18. Bratislava: Alfa 1989, s. 167 – 183.
- IORDANSKAYA, Lidiya – MELCHUK, Igor: On the semantics of Russian causal prepositions. In: *Linguistic Journal*. 1996, No. 2, p. 162–211. (in Russian)
- KŘÍŽKOVÁ, Helena: Adverbiální determinace s významem časovým a příčinným. In: *Slavia*, 1967, roč. 36, s. 507-531.
- KROUPOVÁ, Libuše. Vztah významu gramatického a lexikálního u předložek. In: *Slovo a slovesnost*, 1980, roč. 41, č. 1, s. 49 – 52.
- LEVONTINA, Irina: <Causal prepositions>. In: *New explanatory dictionary of synonyms of the Russian language*. Moscow, 1997, p. 144 – 152. (in Russian)
- LURAGHI, Silvia. Prepositions in Cause expressions. In: *Papers on grammar 2005*, vol. 12, No. 2, p. 609–619.
- LYASHEVSKAYA, Olga – SHAROV, Sergey: *Frequency Dictionary of the Modern Russian Language (on the Materials of the National Corpus of the Russian Language)*. Moscow: Azbukovnik 2009. 1087 p. (in Russian)
- RAJNOCHOVÁ, Natálie. – RUNŠTUKOVÁ, Naděžda. – VAVŘÍN, Martin: *Korpus InterCorp – ruština, verze 13 z 1. 11. 2020. Ústav Českého národního korpusu FF UK, Praha 2020. Dostupný z WWW <http://www.korpus.cz> [cit. 29. 01. 2021].*

ROSEN Alexandr – VAVŘÍN, Martin – ZASINA, Adrian: Korpus InterCorp – čeština, verze 13 z 1. 11. 2020. Ústav Českého národního korpusu FF UK, Praha 2020. Dostupný z WWW <http://www.korpus.cz> [cit. 29. 01. 2021]

Statistika čestiny. Praha, NLN 2009. 215 s.

ŠTÍCHA, František a kol.: Velká akademická gramatika spisovné češtiny. I. Morfologie. Praha: Academia, 2018. 1148 s.

VSEVOLODOVA, Maya – YASHCHENKO, Tatyana: Causal relationships in modern Russian. Moscow: Russkiy yazyk 1988. 208 p. (in Russian)

ZOLOTOVA, Galina: Syntactical Dictionary: a set of elementary units of the Russian syntax. 4th edn. Moscow 2011. (in Russian)