Chapter 14

Polar questions in Czech and Russian: An exploratory corpus investigation

Maria Onoeva & Anna Staňková
Charles University

This study aims to bring new insights into the topic of polar questions in Czech and Russian based on corpus data. What is of particular interest are the complex differences in meaning among the Czech and Russian counterparts of English structures such as *Is Jane coming?*, *Isn't Jane coming?* and *Is Jane not coming?*. We examine the formal and semantic/pragmatic features of polar questions in these two Slavic languages, namely word order, presence and position of negation, presence of question tags, presence of question particles, and their relation to the question's meaning and its bias towards a possible answer. Using authentic data from spoken corpora allowed us to observe some prominent tendencies of polar questions usage.

1 Introduction

Polar questions (PQs) have been widely studied from different points of view in recent years.[1] A number of researchers brought important insights about how the meaning of PQs is affected by negation (e.g. Büring & Gunlogson 2000, Romero & Han 2004, Repp 2013, AnderBois 2019), word order (e.g. Gunlogson 2002), particles (e.g. Sudo 2013, Frana & Rawlins 2019, Gärtner & Gyuris 2022, Gonzalez 2023), intonation or focus (e.g. Gyuris 2019, Rudin 2022, Goodhue 2022), and other phenomena. The goal of this paper is to contribute to the topic from the perspective of Slavic languages, namely Czech and Russian.

[1]The term "question" is related rather to a speech act, whilst "interrogative" is used for syntactic and semantic descriptions. In this paper, however, we stick to the term "polar question" as it is more frequent in the literature.

Maria Onoeva & Anna Staňková. 2025. Polar questions in Czech and Russian: An exploratory corpus investigation. In Berit Gehrke, Denisa Lenertová, Roland Meyer, Daria Seres, Luka Szucsich & Joanna Zaleska (eds.), *Advances in formal Slavic linguistics 2022*, 429–446. Berlin: Language Science Press. DOI: 10.5281/zenodo.15394189

There are two main strategies how to ask a PQ in these languages – overt and intonational (Šimík forthcoming). In Czech, a PQ can be constructed by interrogative word order (Štícha 1995), which involves the finite verb preceding an overt subject, as in (1a). The second strategy is using intonation, either the rise or fall-rise pattern (Daneš et al. 1987, Palková 1994). Thanks to this, declarative sentences can be interpreted as PQs, as shown in (1b). The results of the experiment run by Staňková (2023) showed that the choice between these two strategies (interrogative vs. declarative) can be motivated by the presence of evidential bias (more on bias in Section 2.2).

(1) a. Koupil si Petr auto?
bought REFL Petr car
'Did Petr buy a car?'

b. Petr si koupil auto?
Petr REFL bought car
'Petr bought a car?' (Cz)

In Russian, the overt strategy is to place the particle *li* after the first phonological word as in (2a). Any word can appear with it and then be in the question focus (King 1994). The intonational strategy is shown in (2b). For out-of-the-blue PQs, word order is declarative but the pitch locus (a steep rise and an immediate fall; Q-PEAK by Esipova 2024) is placed at the verb, here it is *vyigrala* 'won', whereas in statements it is usually placed at the most deeply embedded argument (*priz* 'prize' in this case) (Meyer & Mleinek 2006, Rathcke 2006). Schwabe (2004) and Brown & Franks (1995) mention the markedness of *li* in main clauses and its ongoing loss in colloquial Russian. Nevertheless, *li* must be still present in embedded PQs, (2c). Esipova & Korotkova (2024) argue that *li*-PQs simply present two alternatives and thus are true neutral questions, whereas intonation PQs convey pressure to respond.

(2) a. Vyigrala li Daša priz?
won LI Daša prize
'Did Daša win a prize?'

b. Daša vyigrala priz?
Daša won prize
'Did Daša win a prize?'

c. Ja ne znaju, vyigrala li Daša priz.
I not know won LI Daša prize
'I don't know whether Daša won the prize.' (Ru)

In the present study, we looked at PQs in general through the lens of corpus data. Besides the above-described features, Czech and Russian questions could contain various elements which directly influence their meaning, such as indefinites, different particles, tags, negation etc. Due to the limitations of the Russian corpus, prosody was not taken into account. Answers to PQs were also laid aside. It was an exploratory study, in which we addressed the following research questions:

1. What are the formal properties of PQs in real communication?
2. Besides the core interrogative semantics, what semantic/pragmatic implications do PQs have?
3. Are there any correlations between the formal and semantic/pragmatic aspect?

To answer the first research question, we annotated each PQ with respect to its structure. For the second research question, we explored question biases (Büring & Gunlogson 2000, Sudo 2013, Gärtner & Gyuris 2017) and their distribution among Czech and Russian PQs. The third research question was to check if there is any relation between their form and meaning.

The paper is organized as follows. In Section 2, we describe the method of annotation. Section 3 reports on the absolute values of the annotated features and the results of the inter-annotator agreement. In Section 4, we discuss the results. Section 5 concludes the paper.

2 Method

In this section, we describe the method and procedure of the annotation, which were the same for both languages. We used the spoken corpus of the Russian National Corpus (Grišina 2005, Grišina & Savčuk 2009) and the ORTOFON v2 corpus (Kopřivová et al. 2020) of the Czech National Corpus, the latter accessed via the KonText interface (Machálek 2014). Both corpora contain informal everyday conversations with the option to display a limited context around the question. Audio is not available for the spoken part of the Russian corpus, thus intonation was not taken into account.

For each language, a random sample of 500 instances was manually collected. We queried for the question mark and filtered out *wh*-interrogatives. In order to address the first and second research questions, the annotated features were

divided into two groups – formal and semantic/pragmatic. They are described below.[2]

2.1 Formal features

We have already mentioned some of the formal features of PQs, such as the specific word order or usage of question particles. In our sample, we annotated word order with respect to the position of the verb – it was either initial, medial or final. As for particles, for each one we marked their presence ('1' = present, '0' = absent).

Previous research paid attention to negation in PQs because of non-trivial implications it involves (e.g. Ladd 1981, Repp 2013). It was claimed that there are two types of negation – inner (≈ semantic) and outer (≈ pragmatic) – and that they differ in their syntactic and semantic/pragmatic features. Inner negation is interpreted and licenses Negative Polarity Items [NPIs] (Negative Concord Items [NCIs] in Czech and Russian), whereas outer negation does not trigger the negative operator per se and licenses Positive Polarity Items [PPIs] (Romero & Han 2004; cf. Goodhue 2022). Based on these observations, we annotated our data set for the presence of negation as well as certain indefinites. For Czech, these were *ně*-indefinites (considered as PPIs) and *ni-*/*žád*-indefinites (considered as NCIs). In Russian, they were *-nibud', -to, koe-* indefinites and *ni-* NCIs.

The last annotated formal feature was the question tag. Tag questions consist of an anchor (the PQ) and a tag. There are different types of tags based on their polarity (e.g. Krifka 2015). The first type agrees in polarity with the PQ (matching tags), the second type is of the opposite polarity than the PQ (reverse tags). Moreover, tags can differ in their intonation patterns (Ladd 1981). In our annotation, we marked their presence, but did not distinguish them any further.

In (3), we provide an example of a Czech PQ annotated from a formal point of view for all the features just mentioned.

(3) Snad ho teďka nebudeš stavět ne?
SNAD him now NEG-will build no
'You're not going to build it now, are you?' (Cz)
Formal annotation: WO: xVx; PRT: 1 *snad*; NEG: 1; INDEF: 0; TAG: 1 *ne*

2.2 Semantic/pragmatic features

After Hamblin (1973), Karttunen (1977), and Groenendijk & Stokhof (1984), the semantic interpretation of questions is represented as a set of their (true or pos-

[2]The complete annotation is available here: https://bit.ly/3xKM9XX

sible) answers. In case of PQs, it can be simplified to $\{p, \neg p\}$ where p is a question radical, e.g. for a question *Is it raining?*, p = *it is raining*.

Aside from that the structure of PQs may indicate a certain favor, or BIAS, towards a particular reply, which is not captured by the set of their possible answers. So far SPEAKER (or EPISTEMIC) and EVIDENTIAL biases are recognized. They usually either support or oppose p. Speaker bias is based on prior and private speaker's beliefs,[3] while evidential bias comes from contextual information available to all interlocutors. Not all PQs are equally biased, it is possible that one of the biases or both are absent. If no bias is present, the question is considered to be neutral. Different combinations of biases and their absence represent BIAS PROFILES of PQs and could be universal to specific question forms or particles (Sudo 2013, Gärtner & Gyuris 2017).

In our annotation, we also distinguished a third type of bias which was related to the speaker's awareness of the answer, and we refer to it as KNOWLEDGE bias. If the speaker knows the answer for sure, the PQ is biased. This type of bias is sort of in between speaker and evidential biases. It typically occurs in exam (Krifka 2011) or surprise echo questions.

To be able to investigate the bias profiles of Czech and Russian PQs, we manually constructed an affirmative prejacent ϕ for each question from their radicals. We performed the following steps to produce it: (i) remove negation if it is present, (4a); (ii) remove particles, question tags and other elements that do not appear in statements as in (4b) and (4c); (iii) if the first or second person pronouns appear, replace them with 'speaker' or 'addressee' as in (4d).

(4) a. Není to kočka?
NEG.is it cat
'Isn't it a cat?' (Cz)
ϕ = It's a cat.

b. A Daník tady bude ne?
and Daník here will.be no
'And Daník is going to be here, isn't he?' (Cz)
ϕ = Daník is going to be here.

c. Neuželi oni tože slyšat kak my rugaemsja?
NEUŽELI they also hear how we argue
'Do they also hear how we argue?' (Ru)
ϕ = They also hear how we argue.

[3] Sudo (2013) suggests that epistemic/speaker bias can also include deontic or bouletic states. Since these are too complex to judge based on written text, we only work with speaker's beliefs.

d. A u tebja pomimo sobački est' eščë kto-nibud'?
and at you besides doggie is else anyone
'Do you have anyone else besides a doggie?' (Ru)
ϕ = The addressee has someone else besides a doggie.

We used affirmative prejacents and not question radicals to decide whether or not speakers had any prior belief. The same applies for evidential bias. With the aid of prejacents, it was easier to judge the type of bias and its value in some controversial cases such as PQs with certain particles or outer negation cases, where it was not clear if the radical was affirmative or not.

Judgments about the biases were based on our intuition as native speakers, which were later compared with judgments from three additional annotators (see Section 3.4). We always annotated the questions in some amount of context in order to detect evidential bias. Due to the limitations of the corpora, we were able to capture only linguistic cues of evidence.

To construct the bias profiles, we assigned each bias one of the three values: '1', '0' and '−1'. The value '1' was assigned if the bias supported the affirmative prejacent. For instance, if the speaker believed that ϕ before posing the question, '1' was assigned to speaker bias. On the contrary, the value '−1' was assigned when the bias went against ϕ or, in the other words, supported that $\neg\phi$. E.g. if in the context there was a cue suggesting that $\neg\phi$, evidential bias for such cases was '−1'. The value '0' was assigned if no bias was detected.

The following examples clarify the annotation of the bias profile. In the context of (5), it is mentioned who is the oldest brother and the youngest, Leonid and Aleksandr Aleksandrovič, respectively. The context supports the prejacent ϕ, hence, the value assigned to evidential bias is '1'. The particle *razve* indicates that the speaker's prior belief was that $\neg\phi$ (Geist & Repp 2023, Korotkova 2023), so Viktor believed that Aleksandr Aleksandrovič was not the youngest. The value assigned to speaker bias is '−1'. Since it is clear from the context in (5) that the speaker now knows that ϕ (the speaker mentions the brothers' age difference explicitly), the value assigned to knowledge bias is '1'.

(5) Context: The addressee says her husband, aged 28 at the time, had two brothers: Leonid, 30, and Aleksandr Aleksandrovič, 27. The speaker asks:
Sp: {Kak / razve} Aleksandr Aleksandrovič mladšij?
how RAZVE Aleksandr Aleksandrovič youngest
'Wait a second, is Aleksandr Aleksandrovič the youngest?' (Ru)
ϕ = Aleksandr Aleksandrovič is the youngest.
Semantic/pragmatic annotation: SPEAKER −1, EVIDENTIAL 1, KNOWLEDGE 1

In (6), the speaker has some prior belief when the addressee leaves because they explicitly say the time of leaving, so the epistemic bias value is '1'. There were no contextual cues, so the evidential and knowledge bias values are '0'.

(6) Context: The speaker promised some sausage to the addressee but did not manage to bring it. They want to do it later and check when the addressee is available.

Sp: Ty pojedeš kolem vosmý nějak?
you leave around eight somehow
'Are you leaving around eight?' (Cz)
ϕ = The addressee is leaving around eight.

Semantic/pragmatic annotation: SPEAKER 1, EVIDENTIAL 0, KNOWLEDGE 0

In (7), there is an example of evidential bias only. The speaker bias value is '0' because the speaker had no prior belief about the prejacent and guesses the number from the context. Since they do not know for sure how many cars were there, the value assigned to knowledge bias is '0'.

(7) Context: The addressee lists how many people were with them on a trip. The speaker assumes the following from the provided number of people:

Sp: Vy jste jeli třema autama?
you AUX went tree cars
'Did you guys travel in three cars?' (Cz)
ϕ = The addressee and the group travelled in 3 cars.

Semantic/pragmatic annotation: SPEAKER 0, EVIDENTIAL 1, KNOWLEDGE 0

The example in (8) was annotated as carrying knowledge bias only. There is no prior belief about the cat being in the speaker's spot and no linguistic evidence of that.

(8) Context: The speaker sees their cat Ryžik sitting in the speaker's spot.

Sp: Ryžik ty čto moë mesto zanjal?
Ryžik you what my spot taken
'Ryžik, have you taken my spot?' (Ru)
ϕ = Ryžik has taken the speaker's spot.

Semantic/pragmatic annotation: SPEAKER 0, EVIDENTIAL 0, KNOWLEDGE 1

3 Results

In this section, we outline the results of the study. We begin with the absolute values for both the formal and semantic/pragmatic features, then we report the results for Czech and Russian separately. The inter-annotator agreement report concludes the section.

3.1 Overall values

Table 1 summarizes the overall frequencies of the formal features we annotated. From the NEG column follows that, out of 500 PQs, only 89 Czech and 79 Russian were negated. Tag questions were much more frequent in Czech (154) than in Russian (46).

The occurrence numbers of the possible verb positions are in the third column. For Czech, the medial position was the most frequent one (220), followed by initial (150) and final (97). For Russian, we have found fewer PQs with the verb placed at the initial position (53), medial was the second frequent option (148), the most popular was final (237). Some cases were excluded from the final analysis because they were not full sentences (e.g. only subject or object with no predicate).

Table 1: Formal features

	NEG	TAG	VERB POSITION		
	1	1	INI	MED	FIN
Czech	89	154	150	220	97
Russian	79	46	53	148	237

The distribution of the biases in the samples is reported in Table 2. The most striking difference between the languages is present in the column SPEAKER with the value '1' representing speaker bias that ϕ (104 instances for Czech and only 35 for Russian). Aside from that the other bias values were distributed equally in both languages.

In the next subsections, we focus on particular form-meaning correlations in Czech and Russian PQs, respectively. For purposes of the following analyses, we pooled non-null speaker bias and evidential bias and excluded the few exceptional cases where the polarity of speaker/evidential bias was not in accord with the question's polarity.

Table 2: The distribution of the biases in the samples

	SPEAKER			EVIDENTIAL			KNOWLEDGE		
	0	1	-1	0	1	-1	0	1	-1
Czech	379	104	17	353	113	34	446	45	9
Russian	453	35	12	364	109	27	465	31	4

3.2 Czech

The two form-meaning correlations we zoomed in on were: (i) the presence of a TAG and the value of speaker bias (SB), and (ii) verb position and the value of evidential bias (EB). According to the null hypothesis, the variables in the pairs (i) and (ii) are independent of each other. According to the alternative hypothesis, the variables in those two pairs correlate. To test the alternative hypothesis for each pair, we ran two chi-square tests on the Czech data. Since there were two tests run on the data, we used Bonferroni correction (α divided by n, where n is the number of tests) to adjust the alpha level. The adjusted alpha level was 0.025 (0.05 divided by 2).[4] We report the absolute values in the contingency tables 3 and 4. Values expected in the case the null hypothesis is true are in brackets. Now we comment on the two pairs of variables individually.

As mentioned above, tag questions occurred frequently in the Czech data set, and mostly they (the anchor) exhibited declarative word order. Table 3 shows the correlation between tag PQs and speaker bias, i.e. previous beliefs of the speaker; which was statistically significant ($\chi^2(1) = 120.9$; $p < .001$). For non-null speaker bias (= biased PQs), tags were present in 77 cases, even though the expected value by the null hypothesis was 31.7. For null speaker bias, the expected value was 120.3, but tags occurred only 75 times.

Table 3: Tag–belief correlation ($p < .001$)

	NO TAG		TAG	
SB ±1	21	(66.3)	77	(31.7)
SB 0	297	(251.7)	75	(120.3)

[4]We thank an anonymous reviewer for pointing out the inconsistencies in our reporting of the results.

In our sample, initial verb position (V INI) negatively correlates with non-null evidential bias, and this correlation was significant ($\chi^2(1) = 11.8$; $p < .001$). This is apparent from Table 4, where the expected value of V INI in non-null EB PQs is higher (42.6) than the actual value (27). What follows from this is that the interrogative word order is preferred in PQs occurring in neutral context, i.e. those, where evidential bias equals '0'. When it was non-null, the PQ exhibited declarative word order (V NON-INI).

Table 4: V-position–evidence correlation ($p < .001$)

	V NON-INI		V INI	
EB ±1	103	(87.4)	27	(42.6)
EB 0	211	(226.6)	126	(110.4)

As for indefinites, they were present in 59 Czech PQs. Fifty-four of them were of the *ně*-type (PPIs) and the rest were of the *ni-* or *žád*-type (NCIs).

Question particles were very sparse in the Czech data set. There was only one occurrence of *náhodou* and one of *snad*.

3.3 Russian

In the Russian sample, various particles were found. The most frequent was the particle *čto li* and its variations (26). The initial *čto* (occasionally also medial) was also quite frequent in PQs, we have detected 14 cases. The presence of the particle *razve* suggestively correlates with the speaker bias: all 10 cases displayed it, 6 of them carried evidential bias. Out of 143 cases with particles, *li* was found in 6 of them. Only one question had the particle *neuželi*. Thirty-eight PQs with indefinites were discovered, 18 of them contained the *nibud'* series indefinites, 14 *to*, 6 *ni*.

Unlike for Czech, verbs (and non-verbal predicates) were not very often at the initial position. The most frequent was the final position but no significant correlation was found between the verb position and any of the biases.

Due to the low number of speaker bias in the Russian sample, we were unable to draw any conclusive results about the correlations between the form and this type of bias. The same applies to tags. We address it in the discussion.

3.4 Inter-annotator agreement

To test the reliability of our annotation, we recruited three students in order to later conduct the inter-annotator agreement for the semantic/pragmatic features. One Czech and two Russian speakers were paid to perform the same annotation of the bias profiles for 100 random instances from the samples. They were given instructions how to judge the biases and the affirmative prejacents for each PQ.

We compared the annotation of the recruited students with our own. Table 5 summarizes the results, the annotators are in complete agreement if $\kappa = 1$ (Cohen 1960). The agreement for Czech was moderate to substantial. For Russian, it was poor in the first case, slightly better in the second, moderate for evidential bias but still poor for the other two. Potential reasons for it will be discussed in the next section.

Table 5: The agreement for the bias annotation, κ

	Czech	Russian 1	Russian 2
speaker	0.54	0.05	0.24
evidential	0.60	0.16	0.43
knowledge	0.62	0.33	0.06

4 Discussion

In this section we discuss the results and further comment on the process of annotation. Since both Czech and Russian are Slavic languages, we expected them to behave similarly. This assumption was met to some extent. For example, negative PQs were much less frequent than the positive ones in both languages – out of 500 PQs in each language, there were 89 negative PQs in Czech and 79 in Russian. This finding is consistent with previous research which claims that positive PQs are the unmarked way of requesting information. However, the two languages differed substantially in the frequency of tag PQs. There were also differences in the verb position, which was probably connected to the languages' preferred syntactic mechanisms. As for indefinites, their occurrence in our sample was too sparse to draw any generalizations based on them, although it is an issue that we would definitely like to address in the future.

Czech and Russian showed comparable distribution of values of evidential bias and knowledge bias, but they differed in speaker bias. We suppose that speaker

bias was a category too difficult to be objectively judged based on written material, which might have led to this discrepancy between the languages. The overall number of instances with knowledge bias was small for both languages (around 10%). It is perhaps uncommon for speakers to know for sure the answer to the question they ask because it violates Interrogativity principle (Goodhue 2018).[5]

We continue with the discussion for each of the languages separately.

4.1 Czech

In Czech tag PQs, declarative word order correlated with non-null speaker bias, i.e. tag PQs showed a strong tendency to be biased with respect to what the speaker believed. By uttering such a question, the speaker expresses their prior belief, but as it is not definitive, they shift the commitment onto the addressee at the same time. Since our annotation did not go into so much detail, we could not draw any conclusions about the different types of tags, although it would be an interesting follow-up. Our findings agree with previous research, which claims that tag PQs are mostly biased (Bill & Koev 2023).

Additionally, we observed a negative correlation between interrogative word order and null evidential bias. In these PQs, the speaker has no expectation about the possible answer based on public information shared by the participants. If there is, however, a piece of compelling contextual evidence, the declarative word order would be favored. Again we see that declarative PQs tend to carry a bias. Since this is the case, we have a reason to believe that the interrogative word order is the default strategy of forming an unbiased PQ in Czech, supporting previous claims e.g. by Štícha (1995).

The situation is different when it comes to interrogative PQs with negation, as they do come with a bias. By uttering such a question the speaker expresses that they think that one of the alternatives (p or $\neg p$) is possibly true. It seems that this bias is weaker in its meaning than that of English preposed negation, which is claimed to convey that the speaker *believes* that p or $\neg p$. Czech preposed negation is another issue that would deserve a closer look.

4.2 Russian

We had to analyze the results of the semantic/pragmatic features for Russian cautiously, since the inter-annotator agreement was poor for both annotators. The potential explanation for this is the nature of the Russian corpus. Compared

[5] INTERROGATIVITY PRINCIPLE: Ask a question $?p$ only if the context set c does not entail a complete answer to $?p$.

to the Czech corpus, it displays less context around the question; usually there were two additional lines of text. This could drastically influence the judgments of the bias profiles because evidential bias requires as much context as possible. We suppose that the number of PQs biased with respect to what the speaker believes is different from Czech in Table 2 for the same reason, since it is tricky to judge it considering formal features only. However, we have checked the cases where there was at lease some agreement between the annotators.[6] Such PQs usually contained particles or were annotated as carrying no bias. For the cases with particles, agreement is justified since particles are reliable markers of various biases. In no bias agreement cases, we cannot be certain that the PQs are completely neutral due to little context availability and no audio.

Based on our corpus data we cannot conclude that the low number of tags for Russian in Table 1 signals their lack in colloquial speech compared to Czech. The spoken Russian corpus contains various texts from the sixties until the present days (Grišina & Savčuk 2009) and it seems they were annotated differently in the corpus itself. For instance, tags were often separate one-word questions in the older texts, while in the modern ones they were divided by the pause marking slash '/'. Therefore, more investigation is required, preferably with audio.

When it comes to particles, it is not surprising that *li* was not very frequent in the spoken corpus. As mentioned in the introduction, it is quite marked in colloquial Russian or used in truly neutral contexts but it was not completely absent.

The particles *neuželi* and *razve* occurred in our sample and were recently investigated in a series of experiments by Geist & Repp (2023). Their claim is that *neuželi* denotes VERUM, an epistemic operator indicating the speaker's intention to add the proposition in question to the common ground. *Neuželi* is also incompatible with another illocutionary operator FALSUM which is responsible for outer negation interpretation. *Razve* is compatible with both VERUM and FALSUM. Our findings neither support nor dismiss that since the number of the particles (1 and 10, respectfully) was not sufficient to make any constructive judgments; however, all the cases were biased in one way or another.

Čto li (literally translated as 'what whether') was the most frequent particle. Restan (1969) and Dobrovol'skij & Levontina (2014) mention its presumptive and emotional nature, in other words saying it introduces some bias and is infelicitous in out-of-the-blue PQs. Generally, questions with this particle, e.g. *Na ulice dožd' čto li idët?* 'It is raining outside?', are used in contexts when there is an evidence for p and the speaker wants to confirm that p (out of 26 cases, 10 questions had

[6]We thank an anonymous reviewer who recommended to check it.

the value '1' for evidential bias). A private speaker belief that $\neg p$ is possible but not necessary for affirmative PQs but obligatory for negated PQs with the particle. Unlike the mentioned *neuželi* and *razve*, this particle is also available in declaratives and imperatives contributing epistemic modality flavor as non-at-issue meaning (cf. Bernasconi 2023).

5 Conclusion

The goal of our study was to contribute to the empirical investigation of Slavic PQs. By exploring the properties of PQs through corpora, we addressed three research questions concerning their form and meaning. Corpus proved to be a convenient means of investigation, which allowed us to quickly collect authentic language data. We collected a sample of 500 PQs for each language, in which we were able to observe some tendencies, although it was probably not the best way to evaluate meaning shades, such as the biases.

We have run the inter-annotator agreement for semantic/pragmatic features for both languages. The agreement was moderate to substantial for Czech but poor for Russian. We hypothesize that variations in the corpora may account for this discrepancy. To avoid it in future, we suggest to ensure that a corpus shows at least 10 lines of text prior to a query, corpus texts are annotated in a unified fashion and audio is available.

The contribution of our research is mainly empirical. We observed some interesting form-meaning correlations for Czech, for instance, tag PQs tend to express speaker's belief and initial verb PQs mostly do not carry evidential bias. For Russian, we have seen that the intonational strategy is used predominately in spoken language which supports the previous observations. Moreover, we found some different particles that exhibit certain biases, e.g. *neuželi*, *razve* and less studied *čto li*. In future research, we plan to concentrate on particular phenomena, e.g. negation, particles or intonation in Czech and Russian PQs, since we barely touched upon these or did not even consider them in the corpus investigation.

Abbreviations

AUX	auxiliary	PRT	particle
INDEF	indefinite	REFL	reflexive
NEG	negation	WO	word order

Acknowledgments

The study was funded by the Czech Science Foundation (GAČR), project No. 21-31488J. We would like to acknowledge the valuable input (alphabetically) of Kateřina Hrdinková, Roland Meyer, Mariia Razguliaeva, and Radek Šimík. We would like to express our gratitude to the organizers and audience of FDSL 15 for the discussion and insightful comments. We also thank our annotators who helped us during the investigation, the reviewers and the editors of the volume for their valuable feedback.

References

AnderBois, Scott. 2019. Negation, alternatives, and negative polar questions in American English. In Klaus von Heusinger, Edgar Onea & Malte Zimmerman (eds.), *Questions in discourse – Volume 1: Semantics*, 118–171. Leiden: Brill. DOI: 10.1163/9789004378308_004.

Bernasconi, Beatrice. 2023. Polyfunctional particles in spoken Russian: The case of *čto li*. *Journal of Pragmatics* 208. 77–90. DOI: 10.1016/j.pragma.2023.02.008.

Bill, Cory & Todor Koev. 2023. *Question bias from polarity focus*. Ms. ZAS/U. Konstanz. https://www.corybill.com/publications/bill2023question0.pdf.

Brown, Sue & Steven Franks. 1995. Asymmetries in the scope of Russian negation. *Journal of Slavic Linguistics* 3(2). 239–287. https://www.jstor.org/stable/24598950.

Büring, Daniel & Christine Gunlogson. 2000. *Aren't positive and negative polar questions the same?* Ms. USCS/UCLA. https://semanticsarchive.net/Archive/mYwOGNhO/polar_questions.pdf.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1). 37–46. https://doi.org/dghsrr.

Daneš, František, Zdeněk Hlavsa & Miroslav Grepl. 1987. *Mluvnice češtiny*, vol. 3. Prague: Academia.

Dobrovol'skij, Dmitrij & Irina Levontina. 2014. Diskursivnye slova v obščevoprositel'nyx predloženijax: Russko-nemeckie sootvetstvija. In Vladimir P. Selegej, Aleksej V. Bajtin, Vladimir I. Belikov, Igor' M. Boguslavskij, Boris V. Dobrov, Dmitrij O. Dobrovol'skij, Leonid M. Zaxarov, Leonid L. Iomdin, Irina M. Kobozeva, Elena B. Kozerenko, Maksim A. Krongauz, Natalija I. Laufer, Natal'ja V. Lukaševič, Dijana Makkarti, Joakim Nivre, Gennadij S. Osipov, Viktor Raskin, Ėduard Xovi & Sergej A. Šarov (eds.), *Komp'juternaja lingvistika i intellektual'nye texnologii: Po materialam ežegodnoj meždunarodnoj konferencii*

"Dialog" (2014), 138–149. Moskva: Komp'juternaja lingvistika i intellektual'nye texnologii.

Esipova, Masha. 2024. *Prosody across sentence types*. Submitted to "Proceedings of SALT 34". https://lingbuzz.net/lingbuzz/008389.

Esipova, Masha & Natasha Korotkova. 2024. *To li or not to li*. Talk at *Polar Question Meaning[s] Across Languages*. Amsterdam. https://natasha-korotkova.github.io/files/esipova&korotkova2024-poqal-li-not-li.pdf.

Frana, Ilaria & Kyle Rawlins. 2019. Attitudes in discourse: Italian polar questions and the particle *mica*. *Semantics and Pragmatics* 12(16). 1–48. DOI: 10.3765/sp.12.16.

Gärtner, Hans-Martin & Beáta Gyuris. 2017. On delimiting the space of bias profiles for polar interrogatives. *Linguistische Berichte* 251. 293–316.

Gärtner, Hans-Martin & Beáta Gyuris. 2022. On the absence of propositional negation from Hungarian polar e-interrogatives. *Studia Linguistica* 76(3). 661–683. DOI: 10.1111/stul.12191.

Geist, Ljudmila & Sophie Repp. 2023. Responding to negative biased questions in Russian. In Petr Biskup, Marcel Börner (né Guhl), Olav Mueller-Reichau & Iuliia Shcherbina (eds.), *Advances in formal Slavic linguistics 2021* (Open Slavic Linguistics). Berlin: Language Science Press. DOI: 10.5281/zenodo.10123641.

Gonzalez, Aurore. 2023. Interrogative particles in polar questions: The view from Finnish and Turkish. *Glossa: a journal of general linguistics* 8(1). DOI: 10.16995/glossa.6487.

Goodhue, Daniel. 2018. *On asking and answering biased polar questions*. Montreal: McGill University. (Doctoral dissertation). http://id.loc.gov/vocabulary/iso639-2/eng.

Goodhue, Daniel. 2022. Isn't there more than one way to bias a polar question? *Natural Language Semantics* 30. 379–413. https://link.springer.com/article/10.1007/s11050-022-09198-2.

Grišina, Elena. 2005. Ustnaja reč v Nacionaľnom korpuse russkogo jazyka. *Nacionaľnyj korpus russkogo jazyka 2003—2005*. 94–100. https://ruscorpora.ru/new/sbornik2005/07grishina.pdf.

Grišina, Elena & Svetlana Savčuk. 2009. Korpus ustnyx tekstov v NKRJ (RNC): Sostav i struktura. *Nacionaľnyj korpus russkogo jazyka 2006—2008. Novyje rezuľtaty i perspektivy*. 129–149. https://ruscorpora.ru/new/sbornik2008/07.pdf.

Groenendijk, Jeroen Antonius Gerardus & Martin Johan Bastiaan Stokhof. 1984. *Studies on the semantics of questions and the pragmatics of answers*. Amsterdam: UvA. (Doctoral dissertation).

Gunlogson, Christine. 2002. Declarative questions. *Semantics and Linguistic Theory* 12. 124–143. DOI: 10.3765/salt.v12i0.2860.

Gyuris, Beáta. 2019. Thoughts on the semantics and pragmatics of rising declaratives in English and of their Hungarian counterparts. In Gyuris Beáta, Mády Katalin & Recski Gábor (eds.), *K + K = 120: Papers dedicated to L. Kálmán & A. Kornai on the occasion of their 60th birthdays*, 247–280. Budapest: MTA Nyelvtudományi Intézet. DOI: 10.18135/kk120.2019.

Hamblin, Charles L. 1973. Questions in Montague English. *Foundations of Language* 10(1). 41–53. https://www.jstor.org/stable/25000703.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1(1). 3–44. https://www.jstor.org/stable/25000027.

King, Tracy Holloway. 1994. Focus in Russian yes-no questions. *Journal of Slavic Linguistics* 2(1). 92–120. https://www.jstor.org/stable/24599026.

Kopřivová, Marie, Zuzana Laubeová, David Lukeš, Petra Poukarová & Marie Škarpová. 2020. *ORTOFON v2: Korpus neformální mluvené češtiny s víceúrovňovým přepisem*. Praha: Ústav Českého národního korpusu FF UK. https://wiki.korpus.cz/doku.php/en:cnk:ortofon.

Korotkova, Natasha. 2023. Conversational dynamics of Russian questions with *razve*. In Maria Onoeva, Anna Staňková & Radek Šimík (eds.), *Proceedings of Sinn und Bedeutung 27*. Prague: Institute of Czech Language & Linguistic Theory, Faculty of Arts, Charles University.

Krifka, Manfred. 2011. Questions. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2, chap. 66, 1742–1785. Berlin: Walter de Gruyter.

Krifka, Manfred. 2015. Bias in commitment space semantics: Declarative questions, negated questions, and question tags. *Semantics and Linguistic Theory* 25. 328–345.

Ladd, D. Robert. 1981. A first look at the semantics and pragmatics of negative questions and tag questions. In Robert A. Hendrick, Carrie S. Masek & Mary Frances Miller (eds.), *Papers from the seventeenth regional meeting of the Chicago Linguistic Society*, 164–171. http://www.lel.ed.ac.uk/~bob/PAPERS/CLS1981.pdf.

Machálek, Tomáš. 2014. *KonText: Application for working with language corpora*. Praha: Faculty of Arts, Charles University. https://kontext.korpus.cz/.

Meyer, Roland & Ina Mleinek. 2006. How prosody signals force and focus: A study of pitch accents in Russian yes-no questions. *Journal of Pragmatics* 38(10). 1615–1635. DOI: 10.1016/j.pragma.2005.05.011.

Palková, Zdena. 1994. *Fonetika a fonologie češtiny: S obecným úvodem do problematiky oboru*. cs. Google-Books-ID: 1V9iAAAAMAAJ. Karolinum.

Rathcke, Tamara. 2006. A perceptual study on Russian questions and statements. *AIPUK* 37. 51–62. https://www.ipds.uni-kiel.de/pub_exx/aipuk/aipuk_37/37_5_Rathcke.pdf.

Repp, Sophie. 2013. Common ground management: Modal particles, illocutionary negation and verum. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 231–274. Leiden: Brill. DOI: 10.1163/9789004183988_008.

Restan, Per. 1969. *Sintaksis voprositel'nogo predloženija: Obščij vopros.* Oslo: Universitetsforlaget.

Romero, Maribel & Chung-Hye Han. 2004. On negative *yes/no* questions. *Linguistics and Philosophy* 27. 609–658. DOI: 10.1023/B:LING.0000033850.15705.94.

Rudin, Deniz. 2022. Intonational commitments. *Journal of Semantics* 39(2). 339–383. DOI: 10.1093/jos/ffac002.

Schwabe, Kerstin. 2004. The particle *li* and the left periphery of Slavic yes/no interrogatives. In Horst Lohnstein & Susanne Trissler (eds.), *The syntax and semantics of the left periphery*, 385–430. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110912111.385.

Šimík, Radek. Forthcoming. Polar question semantics and bias: Lessons from Slavic/ Czech. In Berit Gehrke & Radek Šimík (eds.), *Topics in the semantics of slavic languages* (Open Slavic Linguistics). Berlin: Language Science Press. https://www.researchgate.net/publication/373494477_Polar_question_semantics_and_bias_Lessons_from_SlavicCzech.

Staňková, Anna. 2023. *The expression of speaker's bias in Czech polar questions.* Prague: Charles University. (MA thesis). https://hdl.handle.net/20.500.11956/182788.

Štícha, František. 1995. Otázky predikátové: Inference, implicitnost a explicitní výrazy ilokučních funkcí I. *Slovo a slovesnost* 56(2). 98–109. http://sas.ujc.cas.cz/archiv.php?lang=en&art=3625.

Sudo, Yasutada. 2013. Biased polar questions in English and Japanese. *Current Research in the Semantics/Pragmatics Interface* 28. 275–295. DOI: 10.1163/9789004183988_009.