

Polar questions in Russian and Czech: An exploratory spoken corpus investigation

Maria Onoeva, Anna Staňková

Charles University

onoevam@ff.cuni.cz, stankova.anna147@gmail.com

The semantics and pragmatics of polar questions (and their responses, not topicalized here) has received considerable attention (Büring & Gunlogson 2000; van Rooy & Šafářová 2003; Romero & Han 2004; Sudo 2013; Krifka 2013; Roelofsen & Farkas 2015; Roelofsen et al. 2013; Esipova 2021), but a detailed empirical investigation of Slavic languages is still missing. We aim to start filling this gap by taking a detailed look at Russian (RU) and Czech (CZ) polar questions (including subtypes like declarative questions) through the lens of spoken corpora. Our ambition is exploration rather than hypothesis testing, although the findings are more or less strongly related to existing generalizations or hypotheses. Our main research questions are: What are the formal properties of polar questions in real communication? What semantic/pragmatic implications (besides the core interrogative semantics) do polar questions have? Are there any correlations between the formal and semantic/pragmatic aspect? Disclaimer: We have not been able to investigate the prosody, whether due to the corpus limitations (no sound available in RU) or due to the limited capacity of the research team.

Corpora and annotation¹ We used the spoken corpus of the Russian National Corpus (Grišina & Savčuk 2009) and the ortofon v2 corpus (Kopřivová et al. 2020) of the Czech National Corpus. A sample of 500 (pseudo)randomly selected instances of polar questions and their surrounding contexts were analyzed; for each question, we extracted (in a manual but pseudo-algorithmic fashion) its affirmative prejacent ϕ (e.g., for *Didn't you drink?* ϕ = 'the addressee drank') and annotated the question for various formal and semantic/pragmatic properties expected to be relevant for question formation and meaning (in Slavic languages). Formal: presence/absence of negation, particles, tags and indefinites; word order wrt verb position. Sem/prag (building on Büring & Gunlogson 2000; van Rooy & Šafářová 2003; Romero & Han 2004; Sudo 2013; Roelofsen et al. 2013, a.o.): speaker's knowledge of the true answer (SK 0 = speaker does not know the answer / 1 = speaker knows that the true answer corresponds to ϕ / -1 = speaker knows that the true answer corresponds to $\neg\phi$), speaker's prior belief (also called epistemic bias), i.e., a belief that the speaker held at least until a short moment prior to uttering the question (SB 0 = speaker has no prior belief about ϕ or $\neg\phi$ / 1 = speaker has a prior belief that ϕ / -1 = speaker has a prior belief that $\neg\phi$), and speaker's expectation about the answer (SE 0 = speaker has no expectation about ϕ or $\neg\phi$ / 1 = speaker expects the answer to be ϕ / -1 = speaker expects the answer to be $\neg\phi$). A prototypical question has SK 0; non-null SK corresponds to rhetorical or echo questions. Questions with non-null SB/SE correspond to various flavors of biased questions; prior belief of ϕ or $\neg\phi$ implies an informative component in the question (epistemic bias); expectations of a particular answer involve reasoning about the hearer's epistemic state and typically correspond to "double-checking" questions (Romero & Han 2004), whereby the speaker is checking the correctness of an inference drawn based on the recent discourse situation (evidential bias). The annotation is partly exemplified in (1), where B is the question to be annotated (with ϕ = 'it is good there') and A provides relevant preceding context; B had no prior belief.

- (1) A [...] vždyť se na to těšíme celej den [...]
PRT REFL on it look.forward.1PL whole day
'[...] we've been looking forward to it all day long'

¹Materials and annotation manual: https://osf.io/xpmuq/?view_only=adb975359c4d4734b9b1a78b7edc0671

B jo je to tam dobrý?
 PRT is it there good
 ‘oh, is it good there?’

[NEG 0; PRT 1 *jo*; WO VSx; SK 0; SB 0; SE 1]

Selected results Table 1 shows overall absolute frequencies of the values of selected variables. While both CZ and RU exhibit a similar proportion of negative questions, there are clear differences in the number of tags (more in CZ, for which we have no explanation) as well as verb position (standalone V not counted): CZ exhibits a bias towards V-initial questions and RU towards verb-final questions (both in line with standard interrogative strategies in these languages). CZ and RU do not differ wrt the values of the *sk* variable, where approx. 90% of all questions were true questions (*sk* 0), and the *se* variable, where nearly 30% of the questions conveyed a speaker expectation. The most dramatic difference is in the values of the *sb* variable, in which CZ has more than twice as many non-null *sb* instances than RU. This could be explained by the RU corpus making less context available than the CZ corpus. The overall number of dedicated question particles (Restan 1972; King 1994; Brown & Franks 1995; McCoy-Rusanova 2017) was very low: *razve* (10), *že* (8), *li* (6), *neuželi* (1); cf. the relatively frequent *čto li* (26).

	NEG	TAG	V POSITION			SK			SB			SE		
	1	1	ini	med	fin	0	1	-1	0	1	-1	0	1	-1
CZ	89	154	158	220	97	446	45	9	379	104	17	353	113	34
RU	79	46	53	148	237	465	31	4	453	35	12	364	109	27

Tab. 1: Overall frequencies

Form–meaning correlations For purposes of the following analyses, we pooled non-null *sb* and *se* and excluded the few exceptional cases where the polarity of *sb/se* was not in accord with the question’s polarity. The presence of a tag correlates significantly with non-null *sb*; see Tab. 2 for CZ (expected frequencies are bracketed; for RU the observed frequencies are very low, but consistent with the pattern in CZ). This is expected given that tags attach to declarative questions, which are in turn prone to expressing beliefs (Gunlogson 2002). (No such correlation is evident for *se*.) Initial verb position negatively correlates with non-null *se*; see Tab. 3 for CZ. This is expected if initial verb in CZ correlates with canonical (non-biased) questions. Despite the low numbers, we see a suggestive correlation between the presence of *razve* in RU and *sb*: all 10 questions with *razve* have a non-null *sb*, proving their biased nature. Further analyses will be reported in the talk.

	TAG 0	TAG 1
SB ±1	21 (66.3)	77 (31.7)
SB 0	297 (251.7)	75 (120.3)

Tab. 2: Tag–belief correlation in CZ ($p < .0001$)

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

Tab. 3: V-position–expectation correlation in CZ ($p < .001$)

Conclusion The corpus of 500 annotated questions per language provides useful baseline frequencies of various formal and semantic/pragmatic kinds of questions and can serve as a starting point for further corpus studies and experiments.

Selected references: Brown & Franks 1995 Asymmetries in the Scope of Russian Negation. *Journal of Slavic Linguistics* 3(2), 239–287. • Büring & Gunlogson 2000 Aren’t positive and negative polar questions the same? Ms. UCLA/UCSC. • Grišina & Savčuk 2009 Korpus ustnykh tekstov v NKRJ: sostav i struktura. *Nacionalnyj korpus russkogo jazyka 2006–2008*, 129–149. • Gunlogson 2002 Declarative questions. In *SALT 12: Proceedings from the 12th Conference on Semantics and Linguistic Theory*, 124–143. • King 1994 Focus in Russian Yes-No

Questions. *Journal of Slavic Linguistics* 2(1), 92-102. • **Kopřivová et al. 2020** ORTOFON v2: Korpus neformální mluvené češtiny s víceúrovňovým přepisem. • **Romero & Han 2004** On negative yes/no questions. *Linguistics and Philosophy* 27(5), 609-658. • **van Rooy & Šafářová 2003**. On polar questions. In *SALT 13: Proceedings from the 13th Conference on Semantics and Linguistic Theory*, 2(36), 202–309. • **Sudo 2013** Biased polar questions in English and Japanese. *Beyond expressives: Explorations in use-conditional meaning*, 275-295. • **Restan 1972** *Sintaksis voprositel'nogo predloženiya: obščij vopros*. Oslo: Universitetsforlaget. • **Roelofsen et al. 2013** Positive and negative questions in discourse. In *Proceedings of Sinn und Bedeutung*, 17, 455-472. • **Roelofsen & Farkas 2015** Polarity particle responses as a window onto the interpretation of questions and assertions. *Language* 91(2), 359-414.