Modus Questions:
Query Models and Frequency
in Russian Text Corpora

Victoria V. Kazakovskaya[1] and Maria V. Khokhlova[2]

[1] Russian Academy of Sciences (Institute for Linguistic Studies),
Tuchkov per. 9, 199053, Saint-Petersburg, Russia
`victory805@mail.ru`
[2] Saint-Petersburg State University,
Universitetskaya nab. 11, 199034, Saint-Petersburg, Russia
`khokhlova.marie@gmail.com`

Abstract. The paper deals with the analysis of modus questions used in dialogues of native Russian speakers, discusses their quantitative properties and characteristics. The research focuses on the development of models describing these questions based on the Russian National Corpus and a newspaper corpus. The results obtained can be applied in various fields of natural language processing, e.g. dialogue systems.

Keywords: text corpora, dialogue, modus questions, Russian

1   Introduction

Modus questions (*MQs*) are interpreted as being question constructions appealing to the intentions of the addressee – their opinion, knowledge, evaluation or explanation. In this sense they are paramount for assimilation of the world of the subjective (the so-called theory of mind) [1]. Propositional attitudes and proposition of modus were studied in many works (see for example [2,3,4,5]). The scope of our research is limited to the analysis of models describing modus questions in Russian texts, particularly the quantitative properties of these types of questions in dialogues and the characteristics of the reactions (replies) given.

Prototypical MQs are represented by constructions containing explicit modus – modus frame[3]: e.g.

1. *Как ты думаешь (вы думаете) / полагаешь (вы полагаете) / считаешь (вы считаете)...?* 'What/How do you: SG or PL think / suppose etc. ?';
2. *Ты думаешь (вы думаете) / полагаешь (вы полагаете) / считаешь (вы считаете), что...?* 'Do you: SG or PL think / suppose etc. that...?';

---

[3] We follow Bally's differentiation between modus and dictum [6].

3. *Почему (отчего) ты так думаешь (вы думаете) / полагаешь (вы полагаете) / считаешь (вы считаете)…?* 'Why do you: SG or PL think / suppose etc. that…?'.

Within dialogic units (or discourse sequences) MQs can occupy not only an initiative position (see above 1 – 3), but also a reactive one: e.g.

4. – *Да что же ты, не можешь стукнуть кулаком по столу? – **Почему не могу?** 4a. Ты думаешь, отчего у меня этот синяк под глазом?* (RNC) 'What are you, you cannot bang your fist on the table? **Why cannot I?** You think, why I have this black eye?'.

The size of modus, means of its expression and types of frame pattern vary extensively in utterances. For example, a completely verbalized modus can be the main predicative part of a compound sentence with an explanatory close (4a).

Incomplete (reduced) modus frames are often presented by parenthesis:

5. *Что, **по-вашему**, это может изменить <…>?* (RN) 'What, **in your opinion**, this can change <…>?'.

Being incorporated in the dictum, modus complicates a simple sentence with an analytic (namely, compound nominal) predicate:

6. – *Неужели вы его, правда, считаете величайшим? — спросил меня тот, кого мы в нашем рассказе условно называем Петровым. <…>* (RNC) 'Do you have him, however, considered the greatest? — Asked me the one whom we in our story conventionally call Petrov. <…>'

2   Materials and Methods

The research was carried out on corpora of the Russian language (compiled by S. Sharoff): 1) a subset of the Russian National Corpus (RNC, 116 million tokens) and 2) a Newspaper Corpus (RN, 70 million tokens) [7]. For the comparative analysis we used data from the Russian National Corpus [8].

Lexico-syntactic models [9] can be used for describing patterns involving lemmas or word forms, part-of-speech tags, characters, and other attributes in an annotated corpus. While writing lexico-syntactic models we used regular expressions and query language IMS Corpus Workbench.

The search of the system is based on morphological annotation combined with lemmata and word forms. For example, the pattern [lemma="как"] []{0,5} [word="считаешь" | word="считаете"] []{0,15} [lemma="?"] describes constructions with the interrogative-relative pronoun *как* ('how') and verb *считать* ('consider/think') in both 2SG or PL forms, with the distance between them being up to 5 words: cf. *как ты считаешь, …* 'how/what do you consider…', *как ты все-таки считаешь, …* 'how do

you ever consider. . . ', *как же ты все-таки считаешь, . . .* 'how do you ever still consider. . . '.

The restriction [ ]{0,15} means that there are up to 15 words between *считаешь (считаете)* 'you consider' and the end of the sentence (a question mark).

The above-discussed models underlying modus questions make up about 1.5% of the total number of question sentences (strictly speaking, words before a question mark) in the Russian National Corpus.

The whole range of modus questions was restricted to the most characteristic models – constructions that include modus frames of mental semantics with the prototypical intentional predicates (*полагать* 'suppose', *считать* 'consider', *думать* 'think') in the second person singular and plural forms inherent to the Russian replication.

3 Analysis Results

The constructions with mental predicates *считать* 'consider' (*считаешь* / *считаете?*) and *думать* 'think' (*думаешь* / *думаете?*) prove to be the most commonly used in the given corpora (see table 1).

In both corpora MQs with verbs in plural form are very frequent, prevailing in RN (about 90%). This fact could be due to the specifics of the corpus, i.e. a high number of interviews (in Russian the polite form is identical to the 2PL one).

The usage of the pronouns *ты* 'you': 2SG and *вы* 'you': 2PL being non-obligatory, nevertheless dominates in both corpora. In RNC there are 127 examples of constructions **Вы (ты)** *полагаете (полагаешь). . . ?* 'Do you: 2SG or 2PL suppose. . . ?' and only 38 cases where the pronoun is omitted[4], while in RN there are 65 MQs with pronouns and only 5 sentences without them:

7. *<. . . >Полагаете, они вас услышали? — Думаю, что нет, к сожалению* (RN). '[Do you] suppose, they heard you? — I think not, unfortunately.'

The search within sentence boundaries can produce some difficulties because a punctuation mark is viewed as a token and has its own tag. Imposing restrictions for search within certain boundaries (e.g. <s> tags for a sentence) can be seen as a solution in this case.

The description of four modus models of question extracted from the corpus data are given below: *КАК-model* 'HOW-model', *НЕУЖЕЛИ-model* 'REALLY-model', *ПОЧЕМУ-model* 'WHY-model' and its version *ОТЧЕГО-model* 'WHY-model'.

---

[4] Russian language belongs to the so-called pro-drop languages.

Table 1: Modus questions in corpora

№	Модель	RN	RNC
1	как полагаешь / полагаете … ?	35	65
2	как думаешь / думаете … ?	445	939
3	как считаешь / считаете … ?	420	374
4	думаешь / думаете … ?	>987	>997
5	полагаешь / полагаете … ?	70	165
6	считаешь / считаете … ?	>993	>997
7	почему … полагаешь / полагаете … ?	3	6
8	почему … думаешь / думаете … ?	15	177
9	почему … считаешь / считаете … ?	36	63
10	неужели … полагаешь / полагаете … ?	0	9
11	неужели … думаешь / думаете … ?	9	132
12	неужели … считаешь / считаете … ?	2	19
13	отчего … полагаешь / полагаете … ?	0	1
14	отчего … думаешь / думаете … ?	0	3
15	отчего … считаешь / считаете … ?	0	0

3.1 КАК (считаешь / думаешь / полагаешь)-model

The pattern identifying the model in a corpus can be written in the following form: [lemma="как"] []{0,5} [word="считаешь" | word="считаете"] []{0,15} [lemma="\?"].

An interrogative construction in the form of a composite sentence turns out to be the most typical for the KAK-model.

One can distinguish between several kinds of reply for YES /NO-questions based on a partial repetition of either the dictum part of the utterance or its modus part combined with relatives (*да* 'yes', *нет* 'no' etc.). Requests to agree or disagree with the previous information, to clarify it or to make an assumption can be found in the semantics of the reactions. See:

(a) repetition of the dictum (predicate):

     8. – ***Как ты думаешь, нужен*** *будет нам еще Пегий пес или нет? – Нет, не нужен, – опять же совершенно уверенно отвечал Кириск* (RNC). 'Do you think we still need a Spotted Dog or not? – No, it is not needed – replied Kirisk, again, quite confidently.'

(b) repetition of the modus (also with negative particle *не* 'not'):

     9. – ***Как вы думаете***, *закон об ограничении пивной рекламы Госдума поддержит? — Думаю, да* (RNC). 'Do you think the law limiting beer advertising will be supported by the State Duma? – I think so.'

(c) repetition of parts of the modus and dictum:

     10. – ***Как Вы считаете, поднимут*** *«Курск» в этом году или нет? – Считаю, что поднимут* (RN). 'Do you think [they] will raise the 'Kursk' this year or not? – I believe that [they] will raise it.'

(d) avoidance of an answer, clarifying or counter-questions:

11. – ***А вы, Таня, как считаете?*** – *Мне-то какое дело? – дернула Таня плечиком* (RNC). 'And you, Tanya, what do you think? – What do I care? – Tanya answered with a shrug of her shoulders.'

Replies to WH-questions are quite similar to the previous ones containing the requested information (dictum) that can be presented by a frame repeating a mental predicate:

12. – *Срок...* ***Как сам думаешь?*** *(= какой срок?) –* ***Думаю***, *неделя (RNC).* 'The term ... How long do you, yourself, think? (= How long?) – I think, a week.'

3.2   НЕУЖЕЛИ (считаешь / думаешь / полагаешь)-model

The present model can be seen as a modal complicated variant of the KAK-model and is more frequent in RNC compared to RN (160 vs. 11 examples). The pattern is as follows: `[lemma="неужели"] []{0,5} [word="считаешь" | word="считаете"] []{0,15} [lemma="\?"]`.

This question is represented by both complex and simple constructions (the latter is more typical for the predicates *полагать* and *считать*).

The replies include the following items:

(a) repetition of the dictum or its fragments:

13. – *Неужели думаешь, что игре удастся избежать политиза-ции, реваншистского акцента? <...>. — Игре, может, и не удастся, а я в это все лезть не хочу* (RNC). 'Do you really think that the game will be able to avoid politicization, revanchist accent? <...> – The game maybe will not be able to avoid this, I do not want to interfere in it.'

(b) repetition of the modus:

14. – *Мама, неужели ты всерьез считаешь, что он мне пара? – Я ничего не считаю, я только вижу, что он тебя любит.* (RNC) 'Mom, do you seriously believe that we are a couple? —— I believe/consider nothing, I only see that he loves you.'

(c) repetition of parts of the modus and dictum:

15. — *Я здесь родился, я люблю этот город, неужели вы думаете, что я хочу оставить о себе плохую память? — Я просто думаю, что на вас давят интересы не столько эстетического плана, сколько денежного.* (RN) 'I was born here, I love this city, do you really think that I want to leave a bad memory about myself? — I just think that you are under financial influences rather than aesthetic ones.'

(d) avoidance of an answer, clarifying or counter-questions:

16. — *Неужели, как Хрущев, считаете, что они не умеют рисовать? — Хотите байку? Раз в Берлине сидим треплемся с товарищем — галерейщиком <...>.* (RN) 'Do you really think, like Khrushchev, that they do not know how to draw? — Would you like to hear a story? Once, in Berlin, we were sitting chatting with a friend who is a gallery owner <...>.'

3.3 ПОЧЕМУ (считаешь / думаешь / полагаешь)-model

The pattern is described as follows: `[lemma="почему"] []{0,5} [word="счи-таешь" | word="считаете"] []{0,15} [lemma="\?"]`.

    The replies include the following items:

(a) repetition of the dictum (the requested information — the reasoning of the point of view) introduced by the subordinate conjunction потому что:

17. – *Почему вы так думаете? – Потому что очень хорошо его знаю! – с мстительным удовольствием сказала Марьяна* (RNC). 'Why do you think so? — Because I know him very well! — Mariana said, with vindictive pleasure. '

(b) repetition of the modus (mental predicates) that does not contain the reasoning of the opinion as the addressee disagrees:

18. *К. Почему вы так думаете? П. Тут и думать нечего. <Никто из наших деревенских Степшу убить не мог>* (RNC) 'K. Why do you think so? P. There is nothing to think about. No one from our village could (not) kill Stepsha.'

3.4 ОТЧЕГО (считаешь / думаешь / полагаешь)-model

The pattern is described as follows `[lemma="отчего"] []{0,5} [word="пола-гаешь" | word="полагаете"] []{0,15} [lemma="\?"]`.

    This model is a variant of the ПОЧЕМУ-model being the least frequent in the corpora: there are 0 examples in RN, and there are only 4 examples in RNC.

    Like the above-mentioned, the replies partially repeat the modus or exemplify the motivation of the point of view:

19. – *Но отчего вы так думаете? – спросил он. – Да хотя бы оттого, что в конце концов я возвращаюсь в реальный мир, — сказал я* (RNC). 'But why do you think so? – he asked. – If only because, in the end, I go back to the real world – I said.'

4 Conclusion and Further Work

The paper presents preliminary results of the study of interrogative constructions in Russian. The analysis shows that modus questions, although not being very frequent in dialogues of native Russian speakers nevertheless represent a certain trait of modern discourse. The *КАК считаете / думаете-model* proves to be the most prototypical one amongst the above-described models. Constructions appealing to the reasoning of the addressee's point of view as well as stylistically marked ones are less common.

References

1. Kazakovskaja, V.V.: Modusnyje voprosy (k probleme izbytochnosti). In: Acta Linguistica Petropolitana: Trudy Instituta lingvisticheskikh issledovanij RAN. Izbytochnost' v grammaticheskom stroje jazyka. Vol. 6, pp.225–242. Nauka, Sankt-Peterburg (2010)
2. Field, H.: Mental representation. In Block N. (ed) Readings in the Philosophy of Psychology 2, pp.78–114. CUP, Cambridge, Harvard University Press, MA (1978)
3. Gak, V.G.: O logicheskom ischislenii semanticheskikh tipov propozitsional'nykh glagolov. In Propozitsional'nyje predikaty v logicheskom i lingvisticheskom aspekte. pp.38–70. Nauka, Moskva (1987)
4. Jaszczolt, K.M. (ed): The pragmatics of propositional attitude reports. Elsevier Science, Oxford (2000)
5. Larson, R.: The grammar of intentionality. In Peter G., Prayer G. (eds) Logical form and language, pp.228–262. Clarendon press, Oxford (2002)
6. Bally, Ch.: Linguistique générale et linguistique française. Leroux, Paris (1932)
7. Corpora of the Russian language, `http://corpus1.leeds.ac.uk/ruscorpora.html`
8. Russian National Corpus, `http://ruscorpora.ru`
9. Khokhlova, M.V.: Razrabotka grammaticheskogo modulja russkogo jazyka dlja spetsial'noj sistemy obrabotki korpusnykh dannykh. In: Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta. Ser. 9. Filologija, vostokovedenije, zhurnalistika. Vol. 2, pp.162–169. Izdatel'stvo Sankt-Peterburgskogo gosudarstvennogo universiteta, Sankt-Peterburg (2010)