Russian Idioms Formed with Interrogative Pronouns and their Syntactic Properties

Leonid L. Iomdin

Institute for Information Transmission Problems, Russian Academy of Sciences
B. Karetnyj per. 19, GSP-4, Moscow, 127994, Russia
iomdin@iitp.ru

Abstract

The paper discusses idiomatic Russian constructions with interrogative pronouns. Particular attention is given to two distinctly different types of such idioms: 1) constructions in which the interrogative word tends to be at the end, as in мало что (знает) ≈ ‘(knows) few things’, редко когда ≈ ‘on rare occasions’, and 2) constructions opening with an interrogative word, as in какой угодно ≈ ‘whichever you please’, кого попало ≈ ‘whoever comes around’, etc.

Keywords

Syntactic idioms, interrogative pronouns, surface syntactic structure

1 Introductory Remarks

This paper continues the author’s work on the integral description₁ of syntactic phrasemes, or syntactic idioms, of Russian (see Iomdin, 2003, 2005, 2006), i.e. such idiomatic units that are syntactically specific compared to non-idiomatic units.² Analysis of syntactic idioms is not only theoretically important but also relevant for NLP tasks such as machine translation or creation of tagged corpora: in the last few years, the author has been involved in the solution of both classes of tasks (see e.g. Apresjan et al., 2003, 2006).

1.1 K-word Uses in Russian

Most of the attention will be given to syntactic idioms that involve a class of Russian pronouns, namely, the pronouns referred to as wh-words (or, in Alexander Isačenko’s terms, K-words, see Isačenko, 1965). These pronouns fall into three main groups:

1 The model example of integral description of lexical units is the New Explanatory Dictionary of Synonyms by Juri Apresjan, who proposed this principle, and his colleagues (Apresjan et. al., 2004).

2 The term “syntactic phraseme” as understood here was introduced in (Boguslavsky & Iomdin, 1982). It is used, in a slightly different way, in (Melčuk, 1987, 1995). A similar, though not identical, understanding of the term “syntactic idiom” can be found in (Nunberg et al., 1994) and (Jackendoff, 1997). Ray Jackendoff focuses on the fact that a syntactic idiom must have variable parts, such as Russian З-у не до X-a, cf. Мне не до обеда ‘I have more important things to attend to than lunch’. Throughout the paper, I will be using the term “syntactic idiom”, which is more commonly accepted in works written in English.
Leonid L. Iomdin

(a) interrogative pronouns, as in Кто пришел? ‘Who came?’ or Когда он пришел? ‘When did he come?’;

(b) relative pronouns, as in Я знаю, кто пришел ‘I know who came’ or Спроси у него, когда он пришел ‘Ask him when he came’,

and (c) indefinite pronouns, as in

(1) Разве тебя обманывал когда? ‘Have I ever deceived you?’ (Boris Ekimov);

(2) Если кто позвонит – буду завтра с утра ‘If anyone calls, (tell them) I will be there tomorrow morning’ (Dina Rubina);

(3) Жду ли чего? жалею ли о чем? ‘Am I waiting for something? Do I have something to regret?’ (Mikhail Lermontov, translation by Svetlana de Loutchek);

(4) Кто свистит, кто мяучит, кто хнычет ‘Some whistle, some meow, some snivel’ (Osip Mandelstam, Stalin Epigram), etc. (In all examples above and below, K-words and, where possible, their English equivalents are shown in boldface). K-words of the latter type are almost exclusively used in poetic or colloquial style and have no direct equivalents among wh-words in English3. In the literary language, they must be replaced by standard indefinites formed from interrogatives with particles -нибудь, -то etc.; cf. (2) and (2a):

(2a) Если кто-нибудь позвонит, я буду завтра с утра.

1.2 Classes of Indefinites Based on K-words

In addition to the class of indefinite pronouns exemplified in (c), Russian has well over a dozen minor type constructions, or syntactic idioms, which include K-words and, together with them, behave like indefinite pronouns. Some of these syntactic idioms have the K-word at the beginning, as in кто угодно ≈ ‘anyone; whoever you please’, что попадется ≈ ‘anything, whatever comes at hand’, где придет ≈ ‘wherever is possible’, как попало ≈ ‘in whatever way, anyway, harum-scarum’, чей бы то ни было ‘no matter whose’ etc.

Other syntactic idioms have the K-word at the end, as illustrated by cases like чёрт знает что ≈ ‘the devil knows what’, мало кто ≈ ‘few people’, много где ≈ ‘in many places, редко когда ≈ ‘on rare occasions’, абь с кем ≈ ‘with the first person to hand’ etc.

These two types of syntactic idioms are in the focus of this paper. Our main goal is to propose adequate surface syntactic representations for these idioms in the Meaning ⇔ Text paradigm.

In my opinion, even though the semantic content of both types of constructions is largely the same, their syntactic behavior is very different.

To corroborate this statement, I will point out the obvious fact that, despite the relatively free word order of Russian, none of these syntactic idioms allow the movement of the K-word,  

Note, however, that wh-word indefinites are not confined to Russian. They are quite common e.g. in German, as illustrated by questions like Hast du was vergessen? ‘Have you forgotten anything?’ or even affirmative sentences like Ich habe was vergessen ‘I have forgotten something’. In Russian, the latter type of usage is entirely excluded: **Я забыл что (и поэтому вернулся)’ lit. ‘I forgot what (and therefore came back)’.
even in the highly colloquial style of speech. In the first type of idioms, the K-word remains at the beginning (пригласи кого угодно ‘invite whoever you please’ but not *пригласи угодно кого) and, in the second type, the K-word must stay at the end: Мало кто понял объяснение ‘Few people understood the explanation’ but never *Кто мало понял объяснение. Note that, beyond the limits of these idioms, the location of K-words is not so rigid; all types of K-words are in principle able to change their regular position in the colloquial or poetic style, cf. interrogative sentences Кого ты видел? ‘Whom did you see’, Ты кого видел? or even Ты видел-то кого.

The different syntactic behavior of the two types of syntactic idioms under study may be accounted for by the hypothesis that, etymologically, they go back to different syntactic entities. Specifically, idioms beginning with a K-word may be considered as compressions, or fusions, of relative clauses; such a process may be exemplified as follows:

\[(5) \text{пригласи того, кого тебе угодно пригласить} \Rightarrow \text{пригласи того, кого тебе угодно пригласить} \Rightarrow \text{пригласи того, кого тебе угодно пригласить} \]

⇒ пригласи кого угодно.

⇒ invite the one whom you like to invite

⇒ invite the one whom you please

Unlike these, a syntactic idiom that end with a K-word may be viewed as a transformation of a full two-clause sentence that includes a relative clause; cf.

\[(6) \text{Один Бог знает, что он делает} \Rightarrow \text{Он делает один Бог знает что} \Rightarrow \text{Один Бог знает что он делает} \]

⇒ God alone knows what he does

⇒ He does (only) God knows what

At the moment, I am unable to provide a fully convincing proof of this assumption; further research is needed, especially in the area of historical syntax. Still, I believe that the assumption of different syntactic nature of the two types of syntactic idioms may serve as a working hypothesis, which should help us create a satisfactory representation for them.

2 Syntactic Idioms Beginning with a K-Word

I will start by a brief outline of these syntactic idioms. The first part of them can be represented by almost any K-word, namely, что ‘what’, кто ‘who’, когда ‘when’, где ‘where’, куда ‘where, whither’, откуда ‘from where, whence’, как ‘how’, зачем ‘what for’, какой ‘which, what kind of’, чей ‘whose’, который ‘which’, почему ‘why’, отчего ‘why, for what reason’, сколько ‘how much’ and even сколь (an obsolete form of сколько used under specific conditions) and почём ‘at what price’. Notable exceptions to this list are

1) the compound K-word что за loosely synonymous with какой\(^4\); cf. Какой <что за> учебник тебе подойдет? ‘What sort of textbook will suit you?’; Мне подойдет какой угодно учебник ‘Any sort of textbook will suit me’, but not *Мне подойдет что угодно для учебник or *Мне подойдет что за угодно учебник;


\(^4\) Discussion of subtle semantic differences between что за and какой can be found in particular in (Raxilina, 1990) and (Podlesskaya, 2007).
Both exceptions are probably explainable by the fact that the respective syntactic idioms would be “too secondary”, since the K-words что за and что <чего> that should produce these idioms are “secondary” with respect to more neutral synonyms какой и почему.

As for the second parts of these idioms, they are represented by a number of words and word combinations of varied degree of idiomaticity. These include, in particular,

1) the predicative adverb угодно ≈ ‘pleases’ that can be combined with any of the K-words given above (examples range from что угодно ≈ ‘whatever you please’ to сколько угодно ≈ ‘as much as you please’ and сколь угодно ≈ ‘in whatever degree it pleases you’); cf. (7) Интервал времени между делениями не может быть сколь угодно большим ‘The time interval between fissions cannot be arbitrarily large’ (Stanislaw Lem);

2) попало, попадет, ни попадя, попадется and a few other frozen forms of the verbs попасть ≈ ‘get’ and попасться ≈ ‘be caught’. These elements can be combined with almost all of the K-words, cf. Он ест что попало ‘He eats whatever comes around’ or (8):

(8) Смотришь конкурс в Сопоте / И глотаешь пыль, / А кого ни попадя / Пускают в Израиль ‘You watch the contest in Sopot and swallow dust/ While they let anyone who comes around go to Israel’ (Vladimir Vysotsky);

etc. Some of these combinations seem rather artificial or deliberate even though they occur in published literary texts, as in

(9) Стопорили фургон на сколько попало времени ≈ ‘They stopped the van for as long a period of time as they could’ (Oleg Pavlov, Oktyabr Magazine, 1998).

As a matter of fact, idioms formed by different frozen verbal forms listed under 2) behave in a slightly varying way: e.g. syntactic idioms with попало, which is singular past of попасть, are the most neutral, while idioms with ни попадя, an obsolete gerund of the same verb accompanied by a negative particle, are stylistically marked. In addition, ни попадя, unlike other forms of this list, never occurs outside the discussed syntactic idioms. I disregard these differences as they do not affect syntactic representations I am going to propose.

3) the frozen expression бы то ни было, which is stylistically neutral, easily used with any of the K-words and can only be found within these syntactic idioms, cf.

(10) Ты больше знаешь, чем кто бы то ни было ‘You know more than anybody’ (Leo Tolstoy);

(11) Приучайте ребенка засыпать в своей кровати без чьего бы то ни было присутствия ‘Train your baby to sleep in its bed without anyone present’ etc;

4) modal predicative adverbs like надо, нужно, необходимо ≈ ‘necessary’, можно ≈ ‘possible’ etc.; cf.

(12) Брат поговорил с кем надо ≈ ‘My brother spoke with whomever was necessary’.

It should be noted that such syntactic idioms are extremely difficult to identify in NLP tasks as they are homonymous with free utterances forming relative clauses: compare (12) and (13):

(13) Спроси, с кем надо поговорить ‘Ask with whom one needs to speak’;
5) A few verbs representing the meaning of ‘wish’ (хотеть, желать ‘wish’, вздуматься, заблагорассудиться ‘at one’s sweet will’) or ‘requirement’ (понадобиться, следовать ≈ ‘need’, потребоваться ≈ ‘require’, прийтись ≈ ‘be compelled’, полагаться or a frozen participle положено ≈ ‘be due’), etc.; cf.

(14) Ты свободен делать что хочешь ‘You are free to do what you like’ (Nikolai Leskov);

(15) Гулял где хотел, ворчал на кого вздумается ≈ ‘He went wherever he pleased, and grumbled at whoever he thought was worth it’ (Andrey Popov);

(16) Отработает, сколько положено, и уедет ≈ ‘He will work for as long as is required of him, and will leave’ (Vasily Shukshin);

6) A host of idiomatic combinations in which K-words are followed with phrases like Бог на душу положит ≈ ‘by guess and by God’, под руку попадется ≈ ‘one can lay hands on’, Бог пошлет ≈ ‘God will grant’ (lit. ‘God will send’) etc.:

(17) Музыку надописать, как Бог на душу положит ≈ ‘Music should be composed by guess and by God’ (Pyotr Tchaikovsky);

(18) Хватай, что под руку попадется ‘grab what(ever) you can lay your hands upon’ (Ludmila Petrushevskaja);

(19) Я жила и все еще живу в Ницце чем Бог пошлет ‘I have lived and am still living in Nice on whatever God gives me’ (Ivan Bunin),

but they can sporadically be used with all, or almost all, other K-words.

Despite the abundance and diversity of these syntactic idioms, their syntactic behavior appears to be quite similar. I believe that the natural solution for the surface syntactic structure (SyntS) incorporating these idioms is as follows: the K-word is the head of the whole idiom which subordinates the head of the remaining part thereof (in most cases, the only remaining element). As for the K-word itself, it may be subordinated to any element of the sentence, the way normal nouns (for что and кто), adjectives (for какой and чей) or adverbs (for где, когда etc.) do. In other words, such an idiom does not constitute a clause and represents a regular element of the sentence such as subject, object, modifier etc.

To give an example, Fig. 1 shows the SyntS5 for the sentence

(20) Пригласи кого угодно ‘Invite whoever you please’.

Fig. 1. Surface syntactic structure for sentence (20)

5 All screenshots represent the operation of the Russian parser of the ETAP-3 MT system. Details of the system may be found in (Apresjan et al, 2003).
In this structure, the K-word КТО in the accusative case instantiates the 1st completive valence of the verb ПРИГЛАШАТЬ, i.e. it acts as its direct object. The adverbial УГОДНО is subordinated to КТО: this link is represented by the auxiliary syntactic relation.

Fig. 2 represents the structure of a sentence with a similar idiom formed with какой:

(21) Пригласи какого угодно одноклассника ‘Invite whichever classmate you please’

The K-word КАКОЙ is a modifier to the noun ОДНОКЛАССНИК which dominates it via the modificative syntactic relation. The adverbial УГОДНО is subordinated to КАКОЙ in a way similar to that of sentence (20).

It must be emphasized that constructions like (21), together with (11) and (15) above, presented the decisive argument for my preferring the proposed syntactic solution with the K-word as the head to the solution in which the K-word is part of the relative clause. It can hardly be disputed that K-words какой in (21), чей in (11) and кого in (15) can only have an external syntactic head and not one within the respective idiom: there are no lexical elements in these idioms that could possibly dominate the K-words (note e.g. that the intransitive word вздумается in (15) can never dominate кого, a noun in the accusative case).

In a way, the proposed solution entails the view of this class of syntactic idioms as entities more or less equivalent to words. This is of course not true: whilst some of these idioms (those with угодно or попало) have progressed far enough to demonstrate hardly any difference from indefinite pronouns formed with particles –нибудь or –то (with the exception of prosodic features), others are still a long way from words as they retain much of the original SyntS of the expression that produced them. Accordingly, the syntactic structure for

(22) Они говорили на каком Бог на душу положит языке ≈ ‘They spoke in whatever language they could choose by guess and by God’

(taken, amusingly, from a report about a linguistic conference that took place in Crimea, where only Russian and Ukrainian was spoken) may seem unnatural as the finite verb положит depends on the adjective какой and thus belongs to the adjectival phrase (Fig. 3).
However, this structure provides an adequate account of the syntactic idiom under analysis and reflects its important syntactic features (including the ability to stand between the modifier and the noun, something that no relative clause may have). Last but not least, such a structure satisfies the projectivity condition, which to my mind corresponds to its nature.

For reasons of space, I will not discuss here extremely interesting semantic features of this class of idioms. I will confine myself to saying that all of them have a clear semantic constituent of choice among an indefinite set of objects, even though such sets are different for every particular idiom. Profound observations and insights of the semantic differences between the individual idioms of this class can be found in a recent paper (Levontina & Shmelev, 2005).

3 Syntactic Idioms ending with a K-Word

This class of idioms shares most of the semantic properties of the class discussed above. However, as I will try to show below, their syntactic behavior differs drastically from the behavior of idioms beginning with a K-word.

We will start with a brief outline of the material representing this class of idioms. In addition to stylistically expressive idioms like черт знает кто <куда, ... > ‘the devil knows who <where etc.>’ discussed in detail in (Iomdin, 2005), this class includes idioms formed with quantitative adverbs мало ‘few’, много ‘many’, редко and several concessive conjunctions хоть, лишь бы, абы6 and, possibly, a few others. (All these elements can be followed by almost any of the K-words listed above (with the exception of semantically contradictory expressions like *мало сколько ‘few how many’ or *много почем ‘many at what price’) but, for reasons of space we will mostly illustrate the idioms containing the K-word кто ‘who’.

1) Мало кто ‘few people’ and много кто ‘many people’ can be illustrated by sentences like

(23) Мнение большинства мало кому интересно ‘The opinion of the majority is interesting to few people’;

(24) Мало ведь кто держит в уме и в сердце десятитомник “полного” Пушкина ‘Indeed, few people keep in their minds and heart the ten-volume edition of the “complete” Pushkin (NLO magazine, 2006, No. 81).

In the next sentence, маловато ‘somewhat too few’, a derivative form of мало, is used:

(25) Три языка маловато кто на хорошем уровне знает ≈ ‘A bit too few people know three languages at a good level’.

(26) Нужно много с кем познакомиться ≈ ‘One needs to meet with many people’

2) Syntactic idioms formed with редко ‘rarely’ have an interesting semantic property: they combine the idea of fewness with the idea of temporal rarity, so that the sentence

---

6 English equivalents of these conjunctions are not given as they defy translation without sufficient context.

7 Certain syntactic aspects of some of these idioms have recently been considered in an interesting paper (Testelets Y. & E. Bylinina, 2005) which used the notion of sluicing. Our approach is rather different.
(27) Редко кто сходил с поезда, редко кто выглядывал из окон
   Rarely who alighted train rarely who looked out from windows
   ≈ ‘Few people on rare occasions left the train, and few people on rare occasions looked
   out of windows’ (Daniil Granin)

necessarily introduces the dynamics of movement; it would be infelicitous to say something
like *редко кто сидел в этом зале ‘few people on rare occasions were sitting in the hall’.

Similarly, the seemingly redundant idiom in sentence (28) combines the idea of the
insufficient number of places with the rarity of occasions:

(28) Хорошее вино сейчас редко где выпьешь
   Good wine now rarely where you drink
   ≈ ‘There are few places now where you can at times drink good wine’ (Fazil Iskander)

3) Syntactic idioms with conjunctions хоть кто ≈ ‘anyone’, лишь бы кто ≈ ‘anyone,
   however mediocre’, aborto кто (synonymous with лишь бы кто) may be illustrated by
   sentences like

(29) Доверчивое, неожиданное прикосновение молодой, честной души смутит хоть
cого ‘A trustful, unexpected touch of a young and honest soul will confuse anyone’
   (Ivan Turgenev);

(30) Нам лишь бы кто не нужен в идиоты ‘We don’t want just any first comer for the
   idiot’ (Boris Levin, from a humorous poem)

(31) Сразу видно, что не абь кто писал, а хороший и умелый писатель ‘You can see at
   once that it was not just anyone who wrote it but a good skilful writer’ (Mikhail Veller).

Naturally enough, the last two idioms have a strong negative polarity and are mostly used in
sentences with overt or implicit negation.

The challenging semantics of these idioms is matched by their challenging and extraordinary
syntax. I believe that, as was the case with idioms like чёрт знает кто discussed in (Iomdin,
2003): sentences with these idioms cannot be adequately represented by a dependency tree
and should be granted a graph with two top nodes, or two syntactic starts, so that in the
syntactic structure for

(32) Мало кто знает три языка ‘Few people know three languages’

there are two heads – знает and мало, while кто is a predicative dependent on знает,
instantiating the verb’s subject valence, and simultaneously a locutive dependent on мало
reflecting the unity of the idiom.

The reasoning underlying this solution is as follows. Let us assume the contrary and start
constructing a dependency tree for (32). There are two likely candidates for the top node:
мало and знает. Let us first assume that it is мало. In this case, the verb знает should be a
direct dependent of мало. What kind of a link is that? For the finite verb to be a dependent, it
must be a link close to a conjunctonal or intersentential link. This is clearly not the case.
Assume now that the top node is знает. In that case, мало should be a dependent of кто, in
order to account for an indisputable syntactic unity of the phrase. This will entail the fact that
all possible dependents of мало, including sentential adverbs or particles like же or ведь quite common in such sentences will belong to the group of the pronoun. Finally, we will have to admit that constructions like мало ли кто это знает ≈ ‘Is it important who knows this’ (they fall out of the scope of our syntactic idioms but share many of their syntactic properties) will have the whole proposition depending on a pronoun, which normally has no dependents at all.  

Our solution will help avoid these oddities and stretches. Besides, it agrees well with the fact that all idioms having the K-word in the end express two propositions each: sentence (32) asserts a) that there are people who know three languages and b) such people are few. At the same time, the two propositions are syntactically merged into a single sentence. The two-apex syntactic representation appears to be a good approximation of this fact.

There is no doubt that, technically, the solution I propose contradicts the strict definition of the classical surface syntactic structure of MTT, which must be a tree with no exceptions whatsoever. However, the very fact that this definition holds true unless very marginal and challenging constructions come into play testifies to its strength and inspirational power, rather than to its weakness – simply because it shows that the limits of this linguistic concept are very broad – almost as broad as they could possibly be.

Acknowledgements

This work has been funded in part by grants Nos. 05-06-80256 and 05-06-80361 from the Russian Foundation of Basic Research, whose support is greatly appreciated. I am very grateful to Igor Mel'čuk, who read the draft with utmost attention and, despite his disagreement with the main idea, contributed a lot to help the author improve the paper.

Bibliography


Boguslavsky, I & L. Iomdin. 1982. Безусловные обороты и фраземы в толково-комбинаторном словаре. (Unconditional locutions and phrasemes in the explanatory

---

8 For reasons of space, I skip other arguments involving prosodic or communicative aspects of such sentences.
combinatorial dictionary). In: Актуальные вопросы практической реализации систем автоматического перевода. Ч. 2. Москва, Изд-во МГУ, 210–222.


