

КАК ОПИСЫВАТЬ ИМПЛИЦИТНЫЕ ЛОГИКО-СЕМАНТИЧЕСКИЕ ОТНОШЕНИЯ: ОПЫТ СОЗДАНИЯ НАБОРА ДАННЫХ НА РУССКОМ ЯЗЫКЕ

А.А. Гончаров

Федеральный исследовательский центр «Информатика и управление» РАН, Россия, Москва; a.gonch48@gmail.com

П.В. Ярошенко

Федеральный исследовательский центр «Информатика и управление» РАН, Московский государственный университет имени М.В. Ломоносова, Россия, Москва; polina.iaroshenko@yandex.ru

Аннотация: В статье представлен подход к описанию имплицитных логико-семантических отношений и опыт его применения для создания соответствующего набора данных на русском языке. В качестве материала выступают тексты двух параллельных подкорпусов Национального корпуса русского языка — английского и французского. С помощью метода поиска с исключением в параллельных текстах был сформирован список переводных соответствий — пар фрагментов, где в тексте на одном языке показатели искомого логико-семантического отношения имеются, а во фрагменте текста на русском языке показатели этого же отношения, напротив, отсутствуют. Исследование выполнено на примере логико-семантических отношений причины и противопоставления, для которых было сформировано более 1000 аннотированных переводных соответствий. Рассматриваются группы признаков, используемых для описания имплицитных логико-семантических отношений. Признаки иллюстрируются примерами из набора размеченных данных.

Ключевые слова: логико-семантические отношения; имплицитность; разметка данных.

Финансирование: Исследование выполнено за счет гранта Российского научного фонда № 24-28-00527, <https://rscf.ru/project/24-28-00527/>.

HOW TO DESCRIBE IMPLICIT DISCOURSE RELATIONS: THE EXPERIENCE OF CREATING A DATASET IN RUSSIAN

A.A. Goncharov

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Russia, Moscow; a.gonch48@gmail.com

P.V. Iaroshenko

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, Lomonosov Moscow State University, Moscow, Russia; polina.iaroshenko@yandex.ru

Abstract: The article presents an approach to the description of implicit discourse relations (DRs) and the experience of its application to the creation of a corresponding dataset in Russian. The material consists of texts from two parallel subcorpora of the Russian National Corpus: English and French. Using the method of search with exclusion in parallel texts, a list

of translation correspondences (pairs of text spans where the text in one language contains markers of the DR in question, and the text span in Russian, on the contrary, does not contain markers of the DR) was composed. The study was conducted using the example of the DRs of cause and contrast, for which more than 1000 annotated translation correspondences were formed. The groups of features used to describe implicit DRs are examined. The features are illustrated with examples from the dataset.

Keywords: discourse relations; implicitness; data labelling.

Funding: The research was supported by the Russian Science Foundation, project No. 24-28-00527, <https://rscf.ru/en/project/24-28-00527/>.

Introduction

Discourse relations (DRs) are relations that are established between units with propositional content and describe the ways in which these units are related to each other¹. Depending on the perspective, such relations may be referred to as logical-semantic relations, rhetorical relations, or simply coherence relations. **Arguments of the DRs** are defined as text spans corresponding to the units connected by these DRs².

DRs can be either explicit or implicit. **Explicit DRs** are precisely expressed using lexical means, such as connectives or other markers that specify the relation's semantics. **Implicit DRs** are not expressed using any such means. Thus, in the example "*I was late for work because there were traffic jams*" the DR of the cause is explicit (expressed by the connective "because"), and in the example "*I was late for work, there were traffic jams*" it is implicit (not expressed).

Discourse relation recognition is a significant challenge in the field of natural language processing (NLP). On the one hand, determining the presence of logical connections between arguments and, on the other, understanding the semantics of these connections is of practical significance for various NLP applications, such as text generation, summarization, sentiment analysis, machine translation, etc.

The fact whether a relation is explicit or not is a significant factor in determining the difficulty of automating the processing of this data. While in the case of explicit DRs, processing can be largely automated by the presence of a marker, implicit DRs require a more sophisticated approach [Xiang et al. 2024], [Cai et al. 2024]. In [Cai et al. 2024] it is noted

¹ It should be noted that the units in some cases may not have a propositional content.

² To be more precise, the arguments of the DRs should be referred to as the actual units of the content plan, rather than the corresponding units of the expression plan (text spans), but this, in our view, would significantly complicate both the presentation and comprehension of the material.

that in experiments with explicit DRs researchers were able to achieve high metrics — more than 90% in F1-score and accuracy — quite quickly, whereas for implicit DRs even the most advanced methods can still only provide accuracy rates of about 70%.

To improve the quality of automatic recognition of implicit DRs, labelled data are required, and the annotation is a complex and time-consuming procedure [Xiang, Wang 2023]. The studies on the recognition of implicit DRs have been conducted mainly using English material. Therefore, the collection of labelled data for other languages is of particular importance [Jiang, He 2020].

The purpose of the article is to present an approach to the description of implicit DRs and the experience of its application to create an appropriate dataset on Russian-language material.

1. Source data and methods

The material consists of texts from two parallel subcorpora of the Russian National Corpus: English and French. It is through the use of parallel texts that it is possible to identify examples that are likely to contain instances of the implicit DR in question. For this task, the method of search with exclusion in parallel texts presented in [Goncharov 2023] and tested in [Goncharov, Inkova 2020] is applied. The point of the method is to obtain a list of such pairs of parallel text spans where a text span in one language contains markers of the DR being searched for, while a text span in another language does not contain markers of the same DR. The method enables us to significantly accelerate the process of creating a set of annotated examples of implicit DRs without having to resort to synthetic data: the only alternative we could find would be to manually analyze all corpus.

The pairs of spans containing implicit DRs of the target type are annotated, and the result of the annotation is saved in the database [Goncharov 2024] in the form of annotated translation correspondence (ATC) [Goncharov, Iaroshenko 2024]. The main structural elements of the ATC are: (1) blocks of the original and translated text, necessary and sufficient for analysing the DR in question, where the DR arguments and, if present, its markers should be highlighted; (2) feature sets of each text block; (3) a set of translation correspondence features. The proposed annotation scheme allows for a significantly greater number of features to be covered than the annotation scheme used in the most commonly employed approach by the authors of the Penn Discourse Treebank (PDTB) to date [Prasad et

al. 2017], [Webber et al. 2019]. In the PDTB approach, when annotating a block of text with a DR, essentially three things are specified: (1) the arguments of DR; (2) the marker of DR (in the case of implicit DRs, also a potential connective that the annotator thinks could be used to express that DR); (3) the type of DR. This corresponds to the first of the three ATC elements (see above) with a single text block feature — DR type.

The study was conducted using the example of DRs of cause and contrast, for which more than 1000 ATCs were formed. Table 1 contains information on the distribution of ATCs by translation directions and DR types (as of 05.12.2024).

Original	Translation	DR	Number of ATCs
Russian	English	cause	168
Russian	English	contrast	65
English	Russian	cause	146
English	Russian	contrast	52
Russian	French	cause	316
Russian	French	contrast	70
French	Russian	cause	135
French	Russian	contrast	50
Total			1002

Table 1: Main quantitative characteristics of the dataset

2. Text block features

The list of features was formed, on the one hand, as a result of the analysis of corpus data and, on the other hand, based on several studies that also considered the problem of defining features relevant for DR recognition [Xiang, Wang 2023], [Bunt, Prasad 2016], [Patterson, Kehler 2013], [Hoek et al. 2018], [Das, Taboada 2018]. The list is not intended to be exhaustive and may be revised in the future. Features can be classified on various bases, but we will be guided by the logic of the ATC structure.

2.1. Features describing one of the DR's arguments

We will consider the first group of features (Table 2).

Feature subgroup	Feature	Presence of a lexical marker	DR of cause	DR of contrast
Syntactic	interrogation		+	+
	exclamation		+	+

Syntactic description of the argument	parenthesis	-	+	-
	complex sentence		+	+
	several sentences		+	+
Verb form	imperative	+	+	+
	transgressive		+	-
Modality	possibility (ontological or deontic)	+	+	+
	wish		+	+
	necessity (ontological or deontic)		+	+
	epistemic possibility or necessity		+	+
Predicate	negation	+	+	+
	mental		+	+
Subjective and evaluative vocabulary	emotion	+	+	+
	evaluation		+	+

Table 2: Features describing one of the DR arguments

The “**Syntactic description of the argument**” subgroup contains features that describe the structure of the argument from a formal point of view and do not imply lexical markers.

The subgroup includes the following features:

- **interrogation**: the argument is a question;
- **exclamation**: the argument is an exclamation;
- **parenthesis**: the argument is a parenthesis;
- **complex sentence**: the argument includes several clauses;
- **several sentences**: the argument extends beyond a single sentence.

The “**Verb form**” subgroup includes the following features:

- **imperative**: the argument contains a verb in the imperative mood;
- **transgressive** (called *дееспучастие*): the argument contains a transgressive.

The “**Modality**” subgroup includes the following features:

- **ontological or deontic possibility** (*мочь* ‘can’, *суметь* ‘to be able to’);
- **wish** (*хотеть* ‘to want’);
- **ontological or deontic necessity** (*надо* ‘to need’, *быть должным* ‘to have to’);
- **epistemic possibility or necessity** (*вероятно* ‘probably’, *должно быть* ‘it must be’).

Each of these features indicates that the argument contains a lexical marker of the corresponding modality. The decision not to distinguish between (1) ontological and deontic modality and (2) epistemic possibility and necessity is due, on the one hand, to the choice of the object of study (DRs) and, on the other hand, to the significant overlap of lexical markers of these modality types, which it was decided not to distinguish, in the Russian language.

The **“Predicate”** subgroup integrates the semantic features of the main clause predicate in the argument:

- **negation**: the argument contains a negation that affects the predicate of the argument's main clause;
- **mental predicate**: the predicate of the main clause of the argument is a mental verb (*думать* ‘to think’).

The **“Subjective and evaluative vocabulary”** subgroup includes the following features:

- **evaluation**: the argument contains a lexical unit that can be categorised as evaluative. For example, evaluation of significance (*важный* ‘important’), complexity (*простой* ‘simple’), degree (*слишком* ‘too much’), etc.
- **emotion**: the argument contains a lexical unit that can be categorised as affective (*радоваться* ‘to be happy’, *грустить* ‘to be sad’). Descriptions of physical manifestations of emotions (*затопать ногами* ‘stomping feet’, *нахмуриться* ‘frowning’, *заплакать* ‘crying’) were also included.

The definition of the boundaries of these semantic categories is a significant linguistic challenge in itself. For the purposes of this study, we have relied our annotation on the RNC semantic tagging system ([RNC], see the categories “evaluation” and “emotion”).

2.2. Features describing both arguments

We will consider the second group of features (Table 3).

Feature subgroup	Feature	Presence of a lexical marker	DR of cause	DR of contrast
Punctuation and placement in the text	colon	-	+	+
	dash		+	+
	semicolon		+	+
	different sentences		+	+
	antonyms		+	+
	synonyms		+	+

Lexical correlations	lexical repetition	+	+	+
	anaphora		+	+
	time marker		-	+
	change of state		-	+
	action — result		-	+
	part — whole		-	+
	set — element		-	+

Table 3. Features describing both DR arguments

The “**Punctuation and placement in the text**” subgroup contains features that describe the structure of both arguments from a formal point of view and do not imply lexical markers.

The subgroup includes the following features:

- colon;
- dash;
- semicolon;

Features are assigned if arguments are separated using one of the given punctuation marks.

- **different sentences**: the arguments are not in the same sentence (each argument does not need to be a complete sentence).

The largest subgroup, “**Lexical correlations**”, contains the following features:

- **anaphora**: one argument contains an anaphor and the other contains its antecedent (only in “content word + pronoun” format; see (1) below);
- **synonyms**: the arguments contain synonyms, which we define quite broadly, periphrases — anaphors, where the anaphor is presented as a content word — are also included (*вещь — предмет* ‘thing — object’; *население — жители* ‘population — inhabitants’; *река — Нева* ‘river — the Neva’);
- **antonyms**: the arguments contain antonyms, different types of antonymy are acceptable, both gradable and complementary antonyms are included (*хорошо — худо* ‘good — bad’; *шутка — всерьёз* ‘a joke — to take something seriously’).
- **lexical repetition**: the arguments contain lexical repetition.
- **time markers**: the arguments contain an indication of time periods (*сегодня — вчера* ‘today — yesterday’);
- **change of state**: one argument describes a situation, the other argument contains an indication that the situation has changed; see (2) below;

- **action — result:** one argument contains an indication of an action, the other argument contains an indication of the result of the action, usually contrary to what was expected (*искать — не найти* ‘to seek — not to find’);
- **part — whole:** one argument contains a meronym, the other a holonym (*лицо — человек* ‘a face — a person’);
- **set — element:** one argument contains an indication of a set, another argument contains an indication of one of the elements of the set. Thus, in the example «Любит собак, не любит кошек» (“He loves dogs, he doesn't like cats”) the set is not mentioned directly, but it is clear that *dogs* and *cats* here act as elements of the set “pets”.

3. Examples from the dataset

We will examine some examples of translation correspondences that have been annotated using the presented list of features.

(1) [Раненых учесть мы не могли]_{arg2} — [их противник забрал с собой]_{arg1}.

[We could not count the wounded]_{arg2}, [**because** the enemy took them away with him]_{arg1}.

Д. Н. Медведев, М. А. Белахова. Это было под Ровно (1948) | Stout hearts (пер. D. Skvirsky; 1961)

In (1) the arguments are related by the DR of cause, which is implicit in the Russian text and expressed by the connective *because* in the English one.

Group of features of individual arguments: argument 2 in both texts is characterised by two features — “negation” (*не; not*) and “modality: possibility” (*могли; could*)³.

Group of features of both arguments: arguments 1 and 2 in both texts have the feature “anaphora” (*раненых — их; wounded — them*). In the Russian text, there is a dash between the arguments, therefore they are assigned the corresponding feature; in the English text the feature is not presented.

(2) Однажды, когда Торопыжки не было дома, Незнайка забрался в автомобиль, который стоял во дворе, и стал дергать за рычаги и нажимать педали. [Сначала у него ничего не получалось]_{arg1}, [потом вдруг машина зафыркала и поехала]_{arg2}.

³ It should be noted that when annotating implicit DRs, the complex character of annotation is crucial. Some features may be significant not by themselves but in combination with other features. Thus, a typical combination is the complex of negation and modality, as in the presented example.

One day when Swifty was out Dunno saw the car standing in the yard. He climbed into it and began pulling levers and pressing pedals. [At first nothing happened]_{arg1}, [but all of a sudden the car gave a sputter and began to move]_{arg2}.

Н. Н. Носов. Приключения Незнайки и его друзей (1954) | The Adventures of Dunno and his Friends (пер. М. Wettlin; 1980)

In (2) the arguments are related by the DR of contrast, which is implicit in the Russian text and expressed by the connective *but* in the English one.

Group of features of individual arguments: argument 1 in both texts is characterised by the feature “negation” (*не*; *nothing*).

Group of features of both arguments: arguments 1 and 2 in both texts have the feature “change of state”. Argument 1 presents the situation — the car does not start and does not move despite Dunno’s efforts; in argument 2 the situation changes — the car starts. Lexical markers for Russian: *не получалось* — *поехала*; for English: *nothing happened* — *began to move*. It should be noted that the feature “change of state” usually implies the presence of several lexical markers necessary to describe the situation. “Change of state” in the Russian text is also supported by the feature “time marker” expressed by a pair of lexical markers *сначала* — *потом* (‘at first — later’).

(3) — Обычное спрашивал, — какой у тебя срок? Я сказал: [просил пять]_{arg1}, [получил десять]_{arg2}.

‘He asked the common questions: — What is your term? I said: [asked for five]_{arg1}, [got ten]_{arg2}’

— Les choses habituelles: combien tu tires, tout ça. Je lui ai répondu: [il en a demandé cinq]_{arg1} [**mais** en a récolté dix]_{arg2}.

В. С. Гроссман. Жизнь и судьба (1959) | Vie et destin (пер. А. Berelowitch; 1980)

In (3), the arguments are related by the DR of contrast, implicit in the Russian text and expressed by the connective *mais* in the French text. The example is remarkable for the feature “action — result” with lexical markers being the ratios *просил пять* — *получил десять* in Russian and *demandé cinq* — *récolté dix* in French. The example also illustrates a typical pattern for the “action — result” feature: the result does not match the expectation.

Conclusion

Thus, a new dataset of more than 1000 ATCs with implicit DRs in Russian has been introduced in the article. Each example in the dataset is provided with a more detailed annotation than in similar datasets created for other languages. The annotation system implies not only the definition of the arguments, semantics, and marker (if present) of this DR but also the formation of a group of features of the text block describing it in terms of structure and semantics. In our opinion, a set of text block features — especially when describing implicit DRs — is one of the most important results of analysing linguistic data. We can assume that further consideration of the features can bring us closer to answering the question of how DRs are understood even in the absence of a marker, and what the role of markers in ensuring text coherence is — to establish DRs or to express DRs that can be inferred from analysing the content of arguments even without a marker.

Список литературы

1. Гончаров А.А. Развитие структуры надкорпусных баз данных // Системы и средства информатики. 2024. Т. 34, № 4. С. 73-84.
2. Гончаров А.А. Поиск с исключением в параллельных текстах // Системы и средства информатики. 2023. Т. 33, № 4. С. 102-114.
3. Гончаров А.А., Ярошенко П.В. Принципы аннотирования имплицитных логико-семантических отношений в параллельных текстах // Информатика и ее применения. 2024. Т. 18, № 3. С. 106-114.
4. Гончаров А.А., Инькова О.Ю. Имплицитные логико-семантические отношения и метод их поиска в параллельных текстах // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конференции «Диалог» (Москва, 17 июня — 20 июня 2020 г.) / Под ред. В. И. Беликов, И. М. Кобозева, Н. В. Лукашевич, В. П. Селегей. Т. 19. М., 2020. С. 310-320.
5. Bunt H., Prasad R. ISO DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations // 12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation: Proceedings. Portorož, Slovenia, 2016. P. 45-54.
6. Cai M., Yang Zh., Jian P. Improving Implicit Discourse Relation Recognition with Semantics Confrontation // Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino: 2024. P. 8828-8839.
7. Das D., Taboada M. RST Signalling Corpus: a corpus of signals of coherence relations // Language Resources and Evaluation, 2018. No. 52, P. 149-184.
8. Hoek J., Zufferey S., Evers-Vermeul J., Sanders T.J.M. The linguistic marking of coherence relations. Interactions between connectives and segment-internal elements // Pragmatics & Cognition, 2018, 25:2. P. 276–309.

9. Jiang D., He J. Tree Framework With BERT Word Embedding for the Recognition of Chinese Implicit Discourse Relations // *IEEE Access*. 2020. Vol. 8. P. 162004-162011.
10. Patterson G., Kehler A. Predicting the Presence of Discourse Connectives // *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 2013. P. 914-923.
11. Prasad R., Webber B., Joshi A. The Penn Discourse Treebank: An Annotated Corpus of Discourse Relations // *Handbook of Linguistic Annotation*. Dordrecht, 2017. P. 1197–1217.
12. Webber B., Prasad R., Lee A., Joshi A. The Penn Discourse Treebank 3.0: Annotation Manual, 2019. [Электронный ресурс]. URL: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf> (дата обращения: 10.03.2025).
13. Xiang W., Liu S., Wang B. Parsing and encoding interactive phrase structure for implicit discourse relation recognition // *Neural Computing and Applications*. 2024. Vol. 36. P. 13783-13797.
14. Xiang W., Wang B. A survey of implicit discourse relation recognition // *ACM Computing Surveys*. 2023. Vol. 55. Iss. 12. 34 p.

References

1. Goncharov A.A. Razvitie struktury nadkorporusnykh baz dannykh [Developing the Structure of Supracorpora Databases]. *Systems and Means of Informatics*, 2024, 34 (5), pp. 73-84. (In Russ.). doi: 10.14357/08696527240406.
2. Goncharov A.A. Poisk s isklyucheniem v parallel'nykh tekstakh [Search with Exclusion in Parallel Texts]. *Systems and Means of Informatics*, 2023, 33(4), pp. 102–114. (In Russ.). doi: 10.14357/08696527230410.
3. Goncharov A.A., Iaroshenko P.V. Printsipy annotirovaniya implitsitnykh logiko-semanticheskikh otnosheniy v parallel'nykh tekstakh [Principles of annotation of implicit logical-semantic relations in parallel texts]. *Informatics and Applications*, 2024, 18(3), pp. 106-114. (In Russ.). doi: 10.14357/19922264240313.
4. Goncharov A.A., Inkova O.Yu. Implitsitnye logiko-semanticheskie otnosheniya i metod ikh poiska v parallel'nykh tekstakh [Implicit Logical-Semantic Relations and a Method of their Identification in Parallel Texts]. *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue"*, 2020, 19, pp. 310-320. (In Russ.). doi: 10.28995/2075-7182-2020-19-310-320.
5. Bunt H., Prasad R. ISO DR-Core (ISO 24617-8): Core concepts for the annotation of discourse relations. *12th Joint ACL-ISO Workshop on Interoperable Semantic Annotation: Proceedings*, 2016. pp. 45-54.
6. Cai M., Yang Zh., Jian P. Improving Implicit Discourse Relation Recognition with Semantics Confrontation. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 2024, pp. 8828-8839.
7. Das D., Taboada M. RST Signalling Corpus: a corpus of signals of coherence relations. *Language Resources and Evaluation*, 2018, 52, pp. 149-184. doi: 10.1007/s10579-017-9383-x.

8. Hoek J., Zufferey S., Evers-Vermeul J., Sanders T.J.M. The linguistic marking of coherence relations. Interactions between connectives and segment-internal elements. *Pragmatics & Cognition*, 2018, 25(2), pp. 276-309. doi: 10.1075/pc.18016.hoe.
9. Jiang D., He J. Tree Framework With BERT Word Embedding for the Recognition of Chinese Implicit Discourse Relations. *IEEE Access*, 2020, 8, pp. 162004-162011. doi: 10.1109/ACCESS.2020.3019500.
10. Patterson G., Kehler A. Predicting the Presence of Discourse Connectives. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 914-923.
11. Prasad R., Webber B., Joshi A. The Penn Discourse Treebank: An Annotated Corpus of Discourse Relations. *Handbook of Linguistic Annotation*. Dordrecht, *Springer Science + Business Media*, 2017, pp. 1197-1217.
12. Webber B., Prasad R., Lee A., Joshi A. *The Penn Discourse Treebank 3.0: Annotation Manual*, 2019. URL: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf> (accessed: 10.03.2025).
13. Xiang W., Liu S., Wang B. Parsing and encoding interactive phrase structure for implicit discourse relation recognition. *Neural Computing and Applications*, 2024, 36, pp. 13783-13797. doi: 10.1007/s00521-024-09709-8.
14. Xiang W., Wang B. A survey of implicit discourse relation recognition. *ACM Computing Surveys*, 2023, 55(12), 34 p. doi: 10.1145/3574134.

Об авторах

Гончаров Александр Анатольевич – научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН; a.gonch48@gmail.com

Ярошенко Полина Владимировна – кандидат филологических наук, научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН; программист Научно-исследовательского вычислительного центра МГУ имени М.В. Ломоносова; polina.iaroshenko@yandex.ru

About the authors

Goncharov Alexander A. – Research Fellow, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences; a.gonch48@gmail.com

Iaroshenko Polina V. – (PhD) Research Fellow, Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences; programmer, Research Computing Center, Lomonosov Moscow State University; polina.iaroshenko@yandex.ru