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

А.А. Гончаров

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

П.В. Ярошенко

Федеральный исследовательский центр «Информатика и управление» РАН, Московский государственный университет имени М.В. Ломоносова, Москва, Россия; 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/.

doi: 10.55959/MSU0130-0075-9-2025-48-06-2

Для цитирования: *Гончаров А.А., Ярошенко П.В.* Как описывать имплицитные логико-семантические отношения: опыт создания набора данных на русском языке // Вестн. Моск. ун-та. Серия 9. Филология. 2025. № 6. С. 24–35.

© Гончаров А.А., Ярошенко П.В., 2025

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

Alexander A. Goncharov

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

Polina 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: This article presents an approach to describing implicit discourse relations (DRs) and its application in creating a corresponding dataset in Russian. The study draws on 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 was compiled (pairs of text fragments where the text in one language contains DR markers, while the Russian text fragment does not contain such markers). The study focuses on the DRs of cause and contrast, for which over 1,000 annotated translation correspondences were formed. The paper examines groups of features used to describe implicit DRs, 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/.

For citation: Goncharov A.A., Iaroshenko P.V. (2025) How to Describe Implicit Discourse Relations: The Experience of Creating a Dataset in Russian. *Lomonosov Philology Journal. Series 9. Philology,* no. 6, pp. 24–35.

Introduction

Discourse relations (DRs) are relations established between units with propositional content, describing the ways in which these units are related to each other[1]. Depending on the perspective, such relations may be referred to as logical-semantic relations, rhetorical relations, or simply coherence relations (these terms differ in their content; for detailed analysis see [Inkova 2019]). **Arguments of DRs** are defined as text spans corresponding to the units connected by these DRs[2].

DRs can be either explicit or implicit. **Explicit DRs** are precisely expressed through lexical means, such as connectives or other markers that

[1] It should be noted that the units in some cases may not have a propositional content; see the classification proposed in [Inkova 2019].

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

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 cause is explicit (expressed by the connective "*because*"), while in "*I was late for work, there were traffic jams*" it is implicit (not expressed).

Discourse relation recognition is a significant challenge in natural language processing (NLP). Determining the presence of logical connections between arguments and understanding the semantics of these connections has practical significance for various NLP applications, such as text generation, summarization, sentiment analysis, machine translation, and others.

Whether a relation is explicit or not is a significant factor in determining the difficulty of automating data processing. While for explicit DRs, processing can be largely automated due to the presence of markers, implicit DRs require a more sophisticated approach [Xiang et al. 2024; Cai et al. 2024]. As noted in [Cai et al. 2024], experiments with explicit DRs achieved high metrics — exceeding 90% in F1-score and accuracy — relatively quickly, whereas even the most advanced methods for implicit DRs can only achieve accuracy rates of approximately 70%.

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

This article aims to present an approach to describing implicit DRs and its application in creating an appropriate dataset based on Russian-language material.

1. Source data and methods

The material comprises texts from two parallel subcorpora of the Russian National Corpus (RNC): English and French. The use of parallel texts enables the identification of examples likely to contain instances of the implicit DR in question. For this task, we apply the method of search with exclusion in parallel texts, described in [Goncharov 2023] and tested in [Goncharov, Inkova 2020]. This method aims to obtain pairs of parallel text fragments where a text fragment in one language contains markers of the DR being searched for, whilst the corresponding fragment in another language does not contain markers of the same DR. The method enables significantly accelerating the process of creating annotated examples of implicit DRs without resorting to synthetic data: the only alternative would be manual analysis of the entire corpus.

The pairs of fragments containing implicit DRs of the target type are annotated, and the annotation results are stored in the database [Gon-

charov 2024] as annotated translation correspondences (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 their markers (if present) should be highlighted[3]; (2) feature sets of each text block; (3) a set of translation correspondence features. The proposed annotation scheme allows for coverage of significantly more features 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, text block annotation with a DR specifies essentially three elements: (1) the arguments of DR; (2) the marker of DR (for implicit DRs, also a potential connective that the annotator believes could express that DR); (3) the type of DR. This corresponds to the first of the three ATC elements (mentioned above) with a single text block feature — DR type.

The study was conducted using the example of DRs of cause and contrast[4], for which more than 1,000 ATCs were formed. Table 1 shows the distribution of ATCs by translation directions and DR types (as of 05.12.2024).

Table 1: Main quantitative characteristics of the dataset

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

[3] In this article, we assume that each DR has two arguments, referred to as "argument 1" and "argument 2". Following [Bunt, Prasad 2016], for asymmetric relations, the argument type is chosen based on the argument semantics (linear order is not considered), while for symmetric relations, conversely, it is based solely on the linear order of arguments. Thus, for the DR of cause (asymmetric DR), "argument 1" will always be the one containing the cause of what is described in "argument 2", while for the DR of contrast (symmetric DR), "argument 1" will be understood as the one that appears linearly earlier in the text than "argument 2". It should be noted that this distinction represents some simplification, as examples from natural language texts show that even relations considered symmetric do not always allow argument reordering without changing the meaning.

[4] A detailed description of the semantics of discourse relations of cause and contrast represents a separate complex research task that is not addressed in this article.

2. Text block features

The feature list was primarily formed based on corpus data analysis. We also drew upon several studies that 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]. It should be noted that this article mainly focuses on describing work with specific corpus material; thus, the list of features is not intended to be exhaustive and may be revised in the future.

2.1. Features describing one of the DR's arguments

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

Table 2: Features describing one of the DR's arguments

Feature subgroup	Feature	Presence of a lexical marker	DR of cause	DR of contrast
Syntactic description of the argument	interrogation	–	+	+
	exclamation		+	+
	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		+	+

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 *деепричастие*): 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 based, firstly, on the choice of the object of study (DRs) and, secondly, on the significant overlap of lexical markers of these modality types in the Russian language.

The **"Predicate"** subgroup integrates the semantic features of the predicate in the argument[5]:

- **negation**: the argument contains a negation that affects the predicate;
- **mental predicate**: the predicate 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, our annotation draws upon 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).

[5] In cases where an argument is a complex sentence, only the predicate of the main clause is considered.

Table 3: Features describing both arguments

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		+	+
Lexical correlations	antonyms	+	+	+
	synonyms		+	+
	lexical repetition		+	+
	anaphora		+	+
	time marker		–	+
	change of state		–	+
	action — result		–	+
	set — element		–	+

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;

These features are assigned if arguments are separated by 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, with different types of antonymy being 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 a pair of time period indications (*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');
- **set — element**: one argument contains an indication of a set, the other argument contains an indication of one of the elements of the set. For example, in a sentence with implicit DR of contrast "*После нас приехал какой-то князь, послал в лавку за* ***шампанским****, нет ни одной* ***бутылки*** *во всем городе, все офицеры выпили*"[6], *шампанским* 'champagne' represents the set, while *бутылки* 'bottle' function as element of the set.

3. Examples from the dataset

Let us 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 (tr. 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*)[7].

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, hence they are as-

[6] From N. V. Gogol. Dead Souls (1835–1852). English translation: "Just after we had left Ponomarev's place, some prince or another arrived in the town, and sent out for some champagne; but not a bottle was there left, for the officers had drunk every one!" Dead Souls (tr. C. J. Hogarth; 1931).

[7] 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. However, the examination of such combinations is beyond the scope of this article.

signed the corresponding feature; in the English text this feature is not present.

(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]$_{arg2}$.

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 (tr. M. 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'). The feature "time marker" is not applicable to the English translation, as the time indication appears only in the first argument (*at first*). Such examples are particularly valuable as they may indicate that a feature characterising only the text block with implicit DR is significant for correctly conveying the relation's semantics: if *потом* was removed from the Russian text, the DR of contrast would be lost.

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

'He asked the usual questions — what's your sentence? 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 (tr. A. 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 mark-

ers 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

This article has introduced a new dataset of more than 1,000 ATCs with implicit DRs in Russian. Each example in the dataset is provided with more detailed annotation than in similar datasets created for other languages. The annotation system involves not only defining the arguments, semantics, and markers (if present) of DRs but also forming a group of features that describe the text block in terms of structure and semantics. In our view, a set of text block features — especially when describing implicit DRs — represents one of the most important outcomes of linguistic data analysis. We can hypothesise that further investigation of these features may bring us closer to understanding how DRs are comprehended even without markers, and what role markers play in ensuring text coherence — whether they establish DRs or merely express relations that can be inferred from argument content analysis even without markers.

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.) / Ред. В. И. Беликов, И. М. Кобозева, Н. В. Лукашевич, В. П. Селегей. Т. 19. М., 2020. С. 310-320.
5. *Инькова О.Ю.* Логико-семантические отношения: проблемы классификации // Связность текста: мереологические логико-семантические отношения. М.: ЯСК, 2019. С. 11–98.
6. *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.
7. *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.
8. *Das D., Taboada M.* RST Signalling Corpus: a corpus of signals of coherence relations // Language Resources and Evaluation, 2018. No. 52, P. 149–184.

9. *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.
10. *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.
11. *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.
12. *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.
13. *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).
14. *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.
15. *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 nadkorpusnykh 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. Inkova O.Yu. *Logiko-semanticheskie otnosheniya: problemy klassifikatsii. Svyaznost' teksta: mereologicheskie logiko-semanticheskie otnosheniya* [Logical-semantic relations: classification problems. Text coherence: Mereological logical semantic relations]. Moscow: LRC Publishing House, 2019. Pp. 11–98. (In Russ.).
6. 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.
7. 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.
8. 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.

9. 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.
10. 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.
11. 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.
12. 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.
13. 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).
14. 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.
15. Xiang W., Wang B. A survey of implicit discourse relation recognition. *ACM Computing Surveys,* 2023, 55(12), 34 p. doi: 10.1145/3574134.

Поступила в редакцию 07.05.2025
Принята к публикации 21.10.2025
Отредактирована 21.11.2025

Received 07.05.2025
Accepted 21.10.2025
Revised 21.11.2025

ОБ АВТОРАХ

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

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

ABOUT THE AUTHORS

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

Polina V. Iaroshenko — PhD, Research Fellow, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences; Research Fellow, Research Computing Center, Lomonosov Moscow State University; polina.iaroshenko@yandex.ru