Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2017"

Moscow, May 31—June 3, 2017

TOWARDS BUILDING A DISCOURSE-ANNOTATED CORPUS OF RUSSIAN

Pisarevskaya D. (dinabpr@gmail.com)[1],
Ananyeva M. (ananyeva@isa.ru)[2],
Kobozeva M. (kobozeva@isa.ru)[2],
Nasedkin A. (kloudsnuff@gmail.com)[3],
Nikiforova S. (son.nik@mail.ru)[3],
Pavlova I. (ispavlovais@gmail.com)[3],
Shelepov A. (alexshelepov1992@gmail.com)[3]

[1]Institute for System Programming of the RAS, Moscow, Russia;
[2]Institute for Systems Analysis FRC CSC RAS, Moscow, Russia;
[3]NRU Higher School of Economics, Moscow, Russia

For many natural language processing tasks (machine translation evaluation, anaphora resolution, information retrieval, etc.) a corpus of texts annotated for discourse structure is essential. As for now, there are no such corpora of written Russian, which stands in the way of developing a range of applications. This paper presents the first steps of constructing a Rhetorical Structure Corpus of the Russian language. Main annotation principles are discussed, as well as the problems that arise and the ways to solve them. Since annotation consistency is often an issue when texts are manually annotated for something as subjective as discourse structure, we specifically focus on the subject of inter-annotator agreement measurement. We also propose a new set of rhetorical relations (modified from the classic Mann & Thompson set), which is more suitable for Russian. We aim to use the corpus for experiments on discourse parsing and believe that the corpus will be of great help to other researchers. The corpus will be made available for public use.

Keywords: rhetorical structure theory, discourse analysis, corpus linguistics, corpus annotation, discourse structure, inter-annotator agreement

ПРИНЦИПЫ РАЗРАБОТКИ ДИСКУРСИВНОГО КОРПУСА РУССКОГО ЯЗЫКА

Писаревская Д. (dinabpr@gmail.com)[1],
Ананьева М. (ananyeva@isa.ru)[2],
Кобозева М. (kobozeva@isa.ru)[2],
Наседкин А. (kloudsnuff@gmail.com)[3],
Никифорова С. (son.nik@mail.ru)[3],
Павлова И. (ispavlovais@gmail.com)[3],
Шелепов А. (alexshelepov1992@gmail.com)[3]

[1]Институт системного программирования РАН, Москва, Россия; [2]Институт системного анализа ФИЦ ИУ РАН, Москва, Россия; [3]НИУ ВШЭ, Москва, Россия

1. Introduction

Discourse analysis is the linguistics level that deals with language units of maximal size (Kibrik, Podlesskaya, 2009, 26). During discourse analysis the text is often represented as a hierarchical tree with its parts connected by various rhetorical relations. Discourse and pragmatics have been considered in Natural Language Processing only in recent years due to the complexity of the approach. Discourse parsing can be used in a wide range of natural language processing tasks, including machine translation evaluation, sentiment analysis, information retrieval, text summarization, information extraction, anaphora resolution, question-answering systems, text classification, etc.—it gives significant performance gain in all these applications, as has been shown by a lot of research.

Creation of corpora with discourse structure has become very popular in recent years because they are then used for developing machine learning algorithms to build automated systems for discourse parsing and analysis. Discourse parsers already exist for several languages, most notably for English (RASTA, SPADE, HILDA, CODRA parsers). However, there are no discourse-annotated corpora for written Russian at the moment, and therefore no possibility of creating an automated discourse parser, and as long as only manual annotation of texts is possible, discourse analysis will not be used in any applications for Russian. That is why it is essential to develop a publicly available discourse-annotated corpus for the Russian language.

In this paper we describe the first steps of building a discourse corpus of Russian: the annotation procedure, including establishing the appropriate set of discourse relations, the process of measuring the inter-annotator agreement, and various challenges we faced along the way.

1.1. Background

There are different approaches to discourse analysis. In Rhetorical Structure Theory (RST) discourse structure amounts to a non-projective tree. Penn Discourse Treebank (PDTB) style is connective-led (PDTB (Webber et al., 2016), TurkishDB (Zeyrek et al., 2013), etc.) or punctuation-led (Chinese Discourse TreeBank (Zhou, Xue, 2015)) and is not presented in a tree form. Models based on cohesive relations (Halliday, Hasan 1976) are also not tree-like. We decided to choose RST to take into consideration not only cohesive markers and discourse cues, but also discourse structure of texts. It is important, for example, for coreference resolution in English—sometimes the most crucial for it is the rhetorical distance and not the linear one, cf. (Loukachevitch et al. 2011).

Therefore, for our corpus we adopt the RST framework (Mann, Thompson, 1988). It represents text as a hierarchy of elementary discourse units (EDUs) and describes relations between them and between bigger parts of text. Some EDUs are more essential and carry more important information (nucleus) than others (satellite). There are two rhetorical relation types: nucleus-satellite and multinuclear. While the first type connects a nucleus and a satellite, the latter includes EDUs that are equally important in the analysed discourse. The set of rhetorical relations can vary; it can include, for instance, such relations as Elaboration, Justify, Contrast, Antithesis, Volitional Result, etc. The rhetorical structure theory claims to be applicable to all languages.

In our work we take into account the existing experience of constructing discourse corpora. There are many RST-annotated corpora of different languages. The most well-known one is the RST Discourse Treebank (Carlson et al., 2003)—an English-language corpus of Wall Street Journal articles (385 articles—176,383 tokens). It is the biggest discourse corpus with a detailed manual. Potsdam Commentary Corpus (Stede, Neumann, 2014) [http://corpus.iingen.unam.mx/rst/manual_en.html] is a German-language corpus that consists of newspaper materials (175 articles—32,000 tokens). CorpusTCC (Pardo et al., 2004) is a corpus of Brazilian Portuguese. It includes 100 introductions (53,000 tokens) to PhD theses. Well-developed are also corpora for other languages: Dutch—Dutch RUG Corpus (van der Vliet et al., 2011), Basque—RST Basque Treebank (Iruskieta et al., 2013), Chinese and Spanish—Chinese/Spanish Treebank as a parallel corpus (Cao et al., 2016), etc. Different sets of rhetorical relations have been created based on the "classic set" (Mann, Thomspon, 1988). For instance, the RST Discourse Treebank makes use of 88 relation types (53 nucleus-satellite and 25 multinuclear relations), the Potsdam Commentary Corpus is based on 31 relation types.

The only existing discourse corpus project for Russian is TED-Multilingual Discourse Treebank. This project contains a parallel corpus of TED talks transcripts for 6 languages, including Russian (along with English, Turkish, European Portuguese, Polish, and German). However, it is based on the principles of the Penn Discourse Treebank annotations framework—discourse connectives as discourse-level predicates with a binary argument structure at a local level (Prasad et al., 2007; Zeyrek et al., 2013)—and not on the RST framework. Besides, this recent effort is still in progress and is not publicly available yet.

The foundation for the project of the Discourse-annotated corpus of Russian was laid by the following works of the research team of the Institute for Systems Analysis, FRC CSC RAS (Ananyeva, Kobozeva, 2016 [1, 2]).

2. Rhetorical Structure Corpus for the Russian Language

The Discourse-annotated corpus of Russian will include texts of different genres (science, popular science, news stories, and analytic journalism). The development of the corpus will be continued for 3 years, during which time we are going to annotate more than 100,000 tokens. The corpus will be available for public use. The user will be able to view annotated texts (represented as discourse trees), search for specific relations (or sequences thereof) and word forms, download the annotated texts in XML format.

2.1. Annotation Principles

After conducting extensive research on discourse corpora of other languages, we have developed a detailed annotation manual. As a tool for annotation we have chosen an open-source tool called rstWeb [https://corpling.uis.georgetown.edu/rstweb/info/], which allows to edit a set of relations and change other features if needed.

International experience of discourse annotation demonstrates that due to grammatical differences between languages, an adaptation of the classic RS theory is necessary for almost all of them. That is why in our project we will, among other things, aim to specify the concept of a discourse unit and the set of rhetorical relations for Russian.

Firstly, we have established a preliminary notion of an elementary discourse unit, which, from a syntactic point of view, we take to be roughly equivalent to a clause (similarly to the classic Mann & Thompson approach). However, there are several notable exceptions, such as nominalization constructions with prepositions like *для* ‘for’ and *из-за* ‘because of’ being classified as an EDU and relative clauses with restrictive semantics not being classified as one.

Secondly, we have discussed main annotation principles and created a detailed manual to guide the annotators. It included description of the following 22 relations, which were based on the “classic set” with the specific features of news and scientific texts in Russian taken into account.

- 16 nucleus-satellite (mononuclear) relations: Background, Cause (with subtypes: Volitional Cause and Non-volitional Cause), Evidence, Effect (with subtypes: Volitional Effect and Non-volitional Effect), Condition, Purpose, Concession, Preparation, Conclusion, Elaboration, Antithesis, Solutionhood, Motivation, Evaluation, Attribution (with subtypes: Attribution1 (precise source specification) and Attribution2 (imprecise source specification)), Interpretation.
- 6 multinuclear relations: Contrast, Restatement, Sequence, Joint, Comparison, Same-unit.

We decided to add Preparation and Conclusion to the set due to the genre properties of scientific and analytic texts. We divided Attribution into two subtypes due to the differing level of precision of specifying the information source in news stories.

There are two strategies of annotators’ work in RST analysis (Carlson et al., 2003). An annotator could apply relations to the segments sequentially, from one segment to another, connecting the current node to the previous node (left-to-right). This method is suitable for short texts, such as news reports, but even in such texts there is a risk of overlooking important relations. The other method is more flexible: the annotator segments multiple units simultaneously, then builds discourse sub-trees for each segment, links nearby segments and builds firstly larger subtrees and after that the final tree, linking key parts of the discourse structure (top-down and bottom-up). It is more suitable for big texts. We chose the second method of tagging since it is more intuitive and easier for the annotator.

For the first 3 texts annotators used the set of discourse relations specified above. The texts were of approximately the same length (34, 26 and 38 sentences respectively). All of them were short news articles. The annotators followed the initial guidelines while annotating pilot texts: they segmented the texts and assigned RST relations to the resulting segments. During subsequent discussion it has become clear that this set of relations was not quite convenient for the annotation since some of the relations were extremely hard to differentiate between. Moreover, we have realized that adopting a "classic set" requires further modifications as some relations are probably more obvious and therefore more common in English than in the Russian language.

2.2. Inter-annotator agreement measurement

One of the main problems with RST tagging is the subjectivity of annotators' interpretation: the same text can be annotated in very different ways (Artstein, Poesio, 2008). However, a simple discussion is not enough to establish the level of inter-annotator agreement (IAA). It must be measured quantitatively using a valid and reliable statistic.

Much like in other discourse analysis tasks (Miltsakaki et al., 2004; Eckle-Kohler, 2015), we faced certain challenges with inter-annotator consistency computation. Although the rules of splitting text into EDUs are relatively straightforward, the resulting segmentation is rarely the same for any text segmented by different people. That is why, for example, the Cohen's kappa coefficient is not suitable in our case. The token-based Fleiss' kappa is also not applicable as we deal with units that consist of several tokens. We have finally selected Krippendorff's unitized alpha as a statistic to measure inter-annotator agreement. It operates on whole annotation spans instead of isolated tokens, it can be calculated for any number of annotators, it can be applied to sparse data, and it can process features of different types, including nominal features in our case. As splitting text into EDUs and labeling relations are two separate tasks, the inter-annotator agreement can be measured separately as well. However, Krippendorff's unitized alpha can (and will in our case) be used for both measurements.

The corpus size used for inter-annotator consistency calculation varies from one project to another. Usually it covers about 30 units (Lacy, Riffe, 1996), but we decided to take texts that contain more units so we could check if relation types in the manual are suitable for further work. The total number of EDUs was approximately 190.

The RST tagging by means of rstWeb tool, which is used by annotators, is done in the browser (see Fig. 2), but the system allows to export the result file as an xml-document, which has the following structure:

```xml
<rst>
    <header>
        <relations>
            <rel name="antithesis" type="rst" />
            <rel name="attribution1" type="rst" />
            <rel name="attribution2" type="rst" />
            <rel name="background" type="rst" />
            <rel name="comparison" type="multinuc" />
            <rel name="contrast" type="multinuc" />
            ...
        </relations>
    </header>
    <body>
        &lt;segment id="1" parent="53" relname="same-unit"&gt;Президент Туниса Зин аль-Абидин бен Али,&lt;/segment&gt;
        &lt;segment id="2" parent="50" relname="span"&gt;управлявший страной 23 года,&lt;/segment&gt;
        &lt;segment id="3" parent="52" relname="comparison"&gt;бежал.&lt;/segment&gt;
        &lt;segment id="4" parent="54" relname="span"&gt;Он покинул Тунис,&lt;/segment&gt;
        &lt;segment id="5" parent="4" relname="elaboration"&gt;где бушуют самые массовые за десятилетия протесты.&lt;/segment&gt;
        &lt;segment id="6" parent="5" relname="interpretation"&gt;Тысячи тунисцев вышли на улицы.&lt;/segment&gt;
        ...
        <group id="50" type="span" parent="1" relname="elaboration"/>
        <group id="52" type="multinuc" parent="53" relname="same-unit"/>
        <group id="53" type="multinuc" parent="54" relname="preparation"/>
        ...
    </body>
</rst>
```

Fig. 1. XML structure of an annotated file

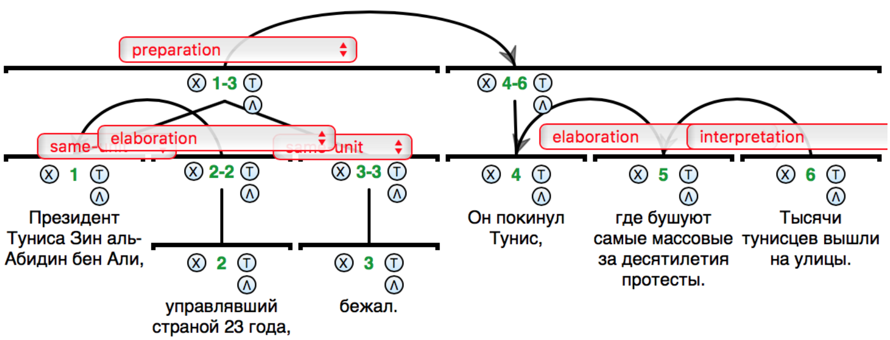

Fig. 2. Annotation in the browser

All the relations used in the scheme are listed in the header of the xml-document.

Each EDU tag includes two ids and a relation type, where "segment id" stands for the id of the EDU, "parent"—for the id of the nucleus in case it is a nucleus-satellite relation and "relname"—for the type of the relation. If the relation is multinuclear, "segment" and "parent" ids both represent the ids of equal by discourse importance EDUs. If the relation type is specified as "span", the EDU is included in a bigger discourse group which is assigned a new id (i.e. the EDUs 4–6 form a bigger group of relations and the EDU 4 as the main nucleus in this group is marked to have a parent with id 54 which is automatically assigned to this group: <segment id="4" parent="54" relname="span">).

Calculation of the IAA coefficient was implemented in Python. Xmltodict 0.10.2 package was used to read and to convert the XML-object of the marked-up text to the Python dictionary. The code used for IAA calculation can be accessed via GitHub [https://github.com/nasedkinav/rst_corpus_rus/blob/master/krippendorffs_alpha.py].

Since the ids of segments and groups may differ in the texts annotated by different people, we have decided to use concatenated text spans to uniquely identify the selected relations since it is the only reliable data between distinct annotations. This format has the additional advantage because it allows to locate identical relations in different parts of text in case of different EDU fragmentation.

During the first iteration of the particular annotator's markup processing, each of the relations trees is traversed in such a way that each node is associated with an ordered by id set of segments of the text, dominated by the node. The "span" relations were not counted during the IAA measurement since this relation plays a structural role in annotation and has no actual meaning. After that, the index of the form {key: value} was produced for all the relations, where the value is the type of the relation, and the key is represented as a string:

- for mononuclear relations: “nuclear: <nuclear_text>, satellite: <satellite_text>”
- for multinuclear relations: “multinuclear: <multinuclear_text>”,

where <nuclear_text>, <satellite_text> and <multinuclear_text> are replaced by correspondent parts of the text. After performing this procedure for each of the annotators, the obtained indices are combined by key <key>, and the list of all the values of the relation, marked by each annotator, is assigned to it. Length of the list can be lower than the number of the annotators when the relation is absent in somebody's markup.

According to (Krippendorff, 2013) we then build the reliability data matrix:

	key_1	key_2	…	key_u	…	key_N
obs_1	$value_{key_1,obs_1}$	$value_{key_2,obs_1}$	…	$value_{key_u,obs_1}$	…	$value_{key_N,obs_1}$
obs_2	$value_{key_1,obs_2}$	$value_{key_2,obs_2}$	…	$value_{key_u,obs_2}$	…	$value_{key_N,obs_2}$
…	…	…	…	…	…	…
obs_m	$value_{key_1,obs_m}$	$value_{key_2,obs_m}$	…	$value_{key_u,obs_m}$	…	$value_{key_N,obs_m}$
Number of coders marked key_u	m_1	m_2	…	m_u	…	m_N

where key_u serves as encoding unit and obs_i stands for particular annotator. Using this matrix, the coincidence matrix within units is calculated (Krippendorff, 2013):

	1	…	k	…	
1	o_{11}	…	o_{1k}	…	n_1
…	…	…	…	…	…
c	o_{c1}	…	o_{ck}	…	$n_c = \sum_k o_{ck}$
…	…	…	…	…	…
	n_1	…	n_k	…	$n = \sum_c \sum_k o_{ck}$

where k, c are concrete relation types and

$$o_{ck} = \sum_u \frac{number\ of\ c - k\ pairs\ in\ unit\ key_u}{m_u - 1},$$

where u is the encoded unit (key_u), m_u is the number of annotators who have marked up this unit.

The final calculation of the coefficient can be done in the following way:

$$\alpha_{nominal} = 1 - \frac{D_o}{D_e} = \frac{A_o - A_e}{1 - A_e} = \frac{(n-1)\sum_c o_{cc} - \sum_c n_c (n_c - 1)}{n(n-1) - \sum_c n_c (n_c - 1)}$$

Fig. 3. Coefficient calculation

We have measured the IAA coefficient for each of three texts and the coefficients for the texts were 0.2792, 0.3173 and 0.4965 respectively. We suppose the third text has the higher IAA coefficient due to the easier and more obvious discourse structure.

The acceptable level of Krippendorff's unitized alpha coefficient for our task would be approximately 0.8 and our results for every text were much lower.

2.3. Initial tree's modification

We have decided to reduce the set of RST relations used for annotation in order to reach the higher IAA coefficient and to minimize the subjectivity of the annotation.

One of the main reasons to exclude particular relations was their high specificity and low frequency of their usage during annotation. Although presence of such relations would not radically affect IAA, reducing the relations' set would make the annotation task easier, and at the same time we would not lose much if we got rid of highly specific and rare relations. If there was always a possibility of replacing some relation with another, more common one, without a great loss in semantic adequacy, it was considered to be an argument in favor of excluding it. The changes that we have accepted after a thorough analysis and much discussion are listed below.

We have decided to exclude from the set of relations

- Motivation, since it is very specific and therefore extremely rare: it was used only 2 times in these three texts (approx. 190 EDUs).
- Antithesis (nucleus-satellite relation), since the only difference between Antithesis and Contrast (multinuclear relation) is that in Antithesis one part should be more important than the other. None of the annotators could establish the relative importance of EDUs in such cases.
- Volitional and Non-Volitional subtypes of Cause and Effect, since in many cases it was impossible to determine whether the actions were intentional or not. However, this distinction might be important for some of the tasks the corpus will be needed for. Those who will use the corpus for this kind of tasks will have the opportunity to substitute Cause/Effect relation with Volitional Cause/Effect or Non-volitional Cause/Effect themselves (as the annotated texts will be available for downloading in an easily changeable XML format).
- Conclusion, because it is quite rare and can be considered a subtype of Restatement, which we decided to use for contexts when the Conclusion relation could be possible.

We have combined in one relation

- Cause and Effect, since the difference between the two lies in determining the nucleus, which is cause in the Cause relation and effect in the Effect relation. Thus, the annotator has to conclude what is more important in two given EDUs: the cause or the effect, which is very subjective.
- Interpretation and Evaluation, since the difference between these relations is very subtle and in order to distinguish between them, one has to determine the degree of objectivity of the evaluation, and that is again very subjective.
- Attribution1 and Attribution2, since the level of precision required for Attribution1 is often unstable and unclear.

All of the above has resulted in a new RST relations tree. The set of relations in Fig. 4 is final and will be used during the rest of the annotation process:

1. **Coherence**
    1.1. Background
    1.2. Elaboration
    1.3. Restatement
    1.4. Interpretation - Evaluation
    1.5. Preparation
    1.6. Solutionhood
2. **Casual-argumentative**
    2.1. Contrastive
        2.1.1. Concession
        2.1.2. Contrast
    2.2. Causal
        2.2.1. Purpose
        2.2.2. Evidence
        2.2.3. Cause-Effect
    2.3. Condition
3. **Structural**
    3.1. Sequence
    3.2. Joint
    3.3. Same-unit
    3.4. Comparison
4. **Attribution**
    4.1. Attribution

Fig. 4. Final set of relations

After modifying the set of discourse relations, three new texts were annotated and the IAA was measured again. The texts were, respectively, 37, 44 and 28 sentences long and all of them were short news articles, same as during the previous IAA measurement. The new IAA coefficients were 0.7768, 0.691 and 0.7615 respectively, which indicates a big leap in the annotation quality. These three texts, annotated in XML format, are available at [https://github.com/nasedkinav/rst_corpus_rus] along with other texts annotated so far. The web interface for the corpus will be created as soon as the appropriate number of texts (and tokens) is reached.

3. Conclusion

By establishing a reliable set of discourse relations we have formed a sound basis for further work. The two iterations of IAA measurement let us believe that using the final relation list will lead to a less biased annotation from now on, which is very important because of the well-known subjectivity of the discourse relations' understanding.

During the rest of the project every text will be annotated by one person and then checked but not annotated by another one. We plan to measure IAA regularly to ensure that the agreement level remains high enough.

After annotating approximately one hundred texts, we plan to conduct several experiments regarding automatic EDUs and discourse relations recognition. Automatic rhetorical structure analysis often relies heavily on determining linguistic discourse markers—connectors that join clauses and sentences into an interconnected piece of text. That is why during the annotation we will also fixate and analyze these markers in order to identify particular words and constructions that indicate discourse relations.

Acknowledgements

We would like to express our sincere gratitude to the advisor of the Discourse-annotated corpus of Russian project, Svetlana Toldova, for the continuous support for the project and related research.

References

1. *Ananyeva M. I., Kobozeva M. B.* (2016), Developing the corpus of Russian texts with markup based on the Rhetorical Structure Theory [Razrabotka korpusa tekstov na russkom yazyke s razmetkoj na osnove teorii ritorecheskikh struktur], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii "Dialog 2016"], Moskva, available at: http://www.dialog-21.ru/media/3460/ananyeva.pdf
2. *Ananyeva M. I., Kobozeva M. B.* (2016), Discourse analysis for natural language processing tasks [Diskursivnyi analiz v zadachakh obrabotki yestestvennogo yazyka], Informatics, management and systems analysis: Proceedings of the IV All-Russian Conference for young scientists [Informatika, upravleniye i sistemnyi analiz: Trudy IV Vserossiiskoi nauchnoi konferencii molodykh uchenykh s mezhdunarodnym uchastiyem], Tver', pp. 138–148, available at: http://www.isa.ru/icsa/images/stories/%D0%A1%D0%B1%D0%BE%D1%80%D0%BD%D0%B8%D0%BA_%D0%A2%D0%BE%D0%BC_1.pdf#page=139
3. *Artstein R., Poesio M.* (2008), Inter-coder agreement for computational linguistics. Computational Linguistics 34(4), pp. 555–596.
4. *Cao S. Y., da Cunha I., Iruskieta M.* (2016), Elaboration of a Spanish-Chinese parallel corpus with translation and language learning purposes, 34th International Conference of the Spanish Society for Applied Linguistics (AESLA), to appear.

5. *Carlson L., Marcu D., Okurowski M. E.* (2003), Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory, Current directions in discourse and dialogue, Kluwer Academic Publishers, pp. 85–112.
6. *Eckle-Kohler J., Kluge R., Gurevych I.* (2015), On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 2236–2242.
7. *Halliday, M. A. K., Hasan R.* (1976), Cohesion in English. London: Longman.
8. *Hayes A. F., Krippendorff K.* (2007), Answering the call for a standard reliability measure for coding data, Communication Methods and Measures Vol. 1, pp. 77–89.
9. *Iruskieta M., Aranzabe M. J., Díaz de Ilarraza A., Gonzalez I., Lersundi M., Lopez de la Calle O.* (2013), The RST Basque TreeBank: an online search interface to check rhetorical relations, IV Workshop RST and Discourse Studies. Fortaleza, Brasil, Outubro 21–23, pp. 40–49.
10. *Kibrik A., Podlesskaya V.* (2009), Stories of dreams: A Corpus-based Research of Russian Oral Discourse [Rasskazy o snovideniyakh: Korpusnoye issledovaniye ustnogo russkogo diskursa], Yazyki slavyanskikh kul'tur, Moskva.
11. *Krippendorff K.* (2013), Computing Krippendorff's Alpha-Reliability, available at: http://web.asc.upenn.edu/usr/krippendorff/mwebreliability5.pdf.
12. *Lacy S., Riffe D.* (1996), Sampling error and selecting intercoder reliability samples for nominal content categories: Sins of omission and commission in mass communication quantitative research, Journalism & Mass Communication Quarterly No 73, pp. 969–973.
13. *Loukachevitch N. V., Dobrov G. B., Kibrik A. A., Khudiakova M. V., Linnik A. S.* (2011), Factors of referential choice: computational modeling [Faktory referencial'nogo vybora: komp'yuternoye modelirovaniye], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2011" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii "Dialog 2011"], Moskva, available at: http://www.dialog-21.ru/media/1446/45.pdf
14. *Mann W. C., Thompson S. A.* (1988), Rhetorical Structure Theory: Toward a Functional Theory of Text Organization, Text 8, 3, 1988, pp. 243–281.
15. *Miltsakaki E., Prasad R., Joshi A., Webber B.* (2004), Annotating discourse connectives and their arguments, Proceedings of the HLT/NAACL Workshop on Frontiers in Corpus Annotation, Boston, Massachusetts, USA, pp. 9–16.
16. *Pardo T. A. S., Nunes M. G. V., Rino L. H. M.* (2004), Dizer: An automatic discourse analyzer for brazilian portuguese, Brazilian Symposium on Artificial Intelligence, Springer Berlin Heidelberg, pp. 224–234.
17. *Prasad R., Miltsakaki E., Dinesh N., Lee A., Joshi A., Robaldo L., Webber B.* (2007). The Penn Discourse Treebank 2.0 Annotation Manual. Technical Report 203, Institute for Research in Cognitive Science, University of Pennsylvania.
18. *Stede M., Neumann A.* (2014), Potsdam Commentary Corpus 2.0: Annotation for Discourse Research. Proc. of LREC, Reykjavik.

19. *Van der Vliet N., Berzlanovich I., Bouma G., Egg M., Redeker G.* (2011), Building a Discourse-Annotated Dutch Text Corpus. Proceedings of the Workshop "Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena", Goettingen, Germany, 23–25 February 2011, pp. 157–171.
20. *Webber B., Prasad R., Lee A., Joshi A.* (2016)., A Discourse-Annotated Corpus of Conjoined VPs. Proc. 10th Linguistics Annotation Workshop, Berlin, pp. 22–31.
21. *Zeyrek D., Demirşahin I., Sevdik Çallı A. B., Çakıcı R.* (2013), Turkish Discourse Bank: Porting a discourse annotation style to a morphologically rich language. Dialogue and Discourse, 4(2), pp. 174–184.
22. *Zhou Y., Xue N.* (2015), The Chinese Discourse TreeBank: A Chinese corpus annotated with discourse relations. Language Resources and Evaluation, pp. 397–431.