

CONTRAST AND COMPARISON RELATIONS IN RST FRAMEWORK: THE CASE OF RUSSIAN¹

Toldova S. (toldova@yandex.ru)¹,
Davydova T. (tdadidik@gmail.com)¹,
Kobozeva M. (kobozeva@isa.ru)²,
Pisarevskaya D. (dinabpr@gmail.com)²

¹NRU Higher School of Economics, Moscow, Russia;

²FRC CSC RAS, Moscow, Russia

The Paper is devoted to a corpus study of the Contrast relation between discourse units in Russian. It is based on the data of the Ru-RSTreebank annotated within the framework of the Rhetorical Structure theory [Mann, Thompson 1988]. The research question is what cue phrases and lexical and grammatical patterns are used to express the Contrast relation as opposed to the Comparison relation. Since the simple connectives such as conjunctions *a* or *no* “but” and others are ambiguous it may be useful to single out specific cues for the Contrast relation and to find other linguistic features that can also help to differentiate Contrast and other relations, such as Comparison. The investigation of cues signalling different types of relations is an important issue for both automatic discourse mining and the theoretical researches of text coherence. We test several hypotheses presented in the reference literature on Russian against corpus data.

Keywords: discourse analysis, rhetorical structure theory, discourse connectives, corpus linguistics, corpus annotation

¹ This paper is partially supported by Russian Foundation for Basic Research (project No. 17-29-07033, 17-07-01477).

ОТНОШЕНИЯ КОНТРАСТА И СРАВНЕНИЯ В ТЕОРИИ РИТОРИЧЕСКОЙ СТРУКТУРЫ НА ПРИМЕРЕ РУССКОГО ЯЗЫКА

Толдова С. (toldova@yandex.ru)¹,
Давыдова Т. (tdadidik@gmail.com)¹,
Кобозева М. (kobozeva@isa.ru)²,
Писаревская Д. (dinabpr@gmail.com)²

¹НИУ ВШЭ, Москва, Россия;

²ФИЦ ИУ РАН, Москва, Россия

Статья посвящена корпусному исследованию отношения Контраста между дискурсивными единицами в русском языке. Используется материал корпуса Ru-RSTreebank, размеченного в рамках Теории риторической структуры [Mann, Thompson 1988]. Простые дискурсивные коннекторы, такие как союзы «а» или «но», неоднозначны и могут маркировать другие отношения, например, Сравнение. Поэтому цель нашего исследования — найти специфические маркеры для Контраста и дополнительные лингвистические параметры, которые помогут дифференцировать Контраст и Сравнение. Мы проверяем несколько гипотез, которые упоминаются в теоретических работах по русскому языку, на материале упомянутого корпуса. Исследование маркеров, указывающих на определенные типы отношений, является важной проблемой как для автоматического анализа дискурса, так и для теоретических исследований связности текста.

Ключевые слова: дискурсивный анализ, теория риторической структуры, дискурсивные маркеры, корпусная лингвистика, корпусная разметка

1. Introduction

It is generally assumed that discourse is not a mere chain of sentences, it is coherent [Hobbs 1985]. Text coherence presupposes that there are relations between text spans. One of the theories modeling discourse as a hierarchical structure built via rhetorical relations between text spans is the Rhetorical Structure theory (RST) [Mann, Thompson 1988]; [Taboada, Mann 2006]. The definitions of relations within RST framework are not based on the explicit linguistic features but formulated in terms of speaker's intentions and its effect on the reader. The aim of our research is to study the contrast relation (Contrast) within RST theory and to establish the repertoire of the linguistic devices (signals) that express it in Russian.

Conjunctions and other cues for Contrast are often ambiguous and unclear:

- (1a) [*S tochki zreniya yazyka zdes' vs'e pravil'no,*] [*a vot s tochki zreniya sootneseniya objektov real'nogo mira—net.*]
 ‘[From the point of view of the language everything is correct here,] [**but** from the point of view of the reference to the real world entities it is not.]’
- (1b) [*Funt i evro slegka ukrepilis’,*] [*a jena slegka upala.*]
 ‘[The pound and the euro strengthened a little bit,] [**while** yen fell slightly.]’

The connective *a* ‘while/but/and’ is used both in (1a) and in (1b), however (1a) is a contrast relation while (1b) is a comparison relation according to the annotation suggested in Ru-RSTreebank (<https://linghub.ru/ru-rstreebank/>, a Russian corpus annotated for rhetorical relations [Pisarevskaya et al. 2017]).

The present research seeks to identify linguistic features that can help to differentiate the contrast and the comparison relations. The results of this study can be helpful for the relation type recognition in the discourse parsing systems.

We use the Ru-RSTreebank as a source of data for Contrast and Comparison. As the data of RU-RSTreebank have shown, a considerable number of discourse connectives (DC) used for Contrast relation are ambiguous as they mark other relations as well. According to the numerous studies on the adversative conjunctions and contrast constructions in Russian [Shvedova 1980]; [Uryson 2004], there is an additional set of features signalling the contrast relation. These include, among others, syntactic parallelism and additional specifiers such as particles of adverbial expressions.

The present work offers a corpus-based analysis of Contrast and Comparison in Russian News and Scientific texts. Its main Hypothesis is that the features suggested in the works on the contrast relations between clauses in Russian are also valid for detecting Contrast within RST framework. The main tasks are to establish the existence of Contrast vs. Comparison-specific connectives and to identify additional features that are specific for these two relations.

For this purpose, we have compiled a set of corresponding discourse units from the Ru-RSTreebank and annotated them manually on different features including discourse markers (DM), additional specifiers and lexical repetitions. In our work we describe the most prominent features that we have identified.

The present paper is organized as follows. A brief description of the framework within which Contrast and Comparison determined is given in 2.1. In 2.2. we discuss the features of conjunctions and other markers of Contrast mentioned in the literature on Russian. The overview of theoretical works on Russian serves as the basis for the hypothesis concerning signals of Contrast and Comparison that are formulated in 2.3. Section 3 describes the data used for the corpus research. In Section 4 we present a survey of the signals enumerated in 2.3 that we have found in our corpus data. Section 5 provides the conclusions.

2. Background

2.1. Contrast and Comparison rhetorical relations in RST framework

According to the RST [Mann, Thompson 1988], a text is organized as a tree whose nodes are text spans (discourse units (DU)). DUs are united into the spans of a higher level if there is a rhetorical relation between them. Relations can be multi-nuclear or mononuclear depending on the types of the spans, e.g.:

[Peter went home]_{nucleus} [because he was tired]_{satellite} vs. [Peter went home]_{nucleus}
[while Tom stayed at work]_{nucleus}.

The two relations under discussion, namely, Contrast and Comparison are multi-nuclear. While Contrast is the relation from the original set given in [Mann, Thompson 1988], Comparison was added later in [Carlson, Marcu 2001]. Below are the definitions of Contrast (a) and Comparison (b) described in [Carlson, Marcu 2001]:

- a. “In CONTRAST, two or more nuclei come in contrast with each other along some dimension. The contrast may happen in only one or few respects, while everything else can remain the same in other respects.”
- b. “In COMPARISON, two textual spans are compared along some dimension, which can be abstract. The relations can convey that some abstract entities that pertain to the comparison relation are similar, different, greater-than, less-than, etc. In the case of a comparison relation, the spans, entities, etc. are not in contrast.”

Thus, the difference between these relations is opaque (cf. “come in contrast” for Contrast—“are different” for Comparison). Moreover, many DC that are used for Contrast are ambiguous between Contrast and Comparison. For example, English *while* is a marker for both—Contrast (2) and Comparison (3) [ibid.]:

- (2) [But the staff at some of those locations will be slashed] [**while** at other the workforce will be increased.]
- (3) [Kellogg’s current share is believed to be slightly under 40%] [**while** General Mills’ share is about 27%.]

The indirect evidence for the fact that Contrast and Comparison are similar to some extent is the fact that for some approaches, e.g. see PDTB, [Prasad et al. 2007: 271] Contrast is a subclass of Comparison (Fig. 1).

Different types of connectives are important cues that help to recognise particular relations and to differentiate them [Danlos 2018]. There are special lexicons compiled for different languages (e.g. French [Roze et al. 2012], Czech [Mirovsky et al. 2017], DiMLex for German [Scheffler, Stede 2016], DiMLex-Eng for English [Das et al. 2018], PDTB based lexicons for French [Laali, Kosseim 2017], and for Portuguese [Mendes et al. 2018]).

In our work, we treat the relations annotated in the Ru-RSTreebank according to their definitions as the starting point and try to find specific connectives and other signals for Contrast and Comparison.

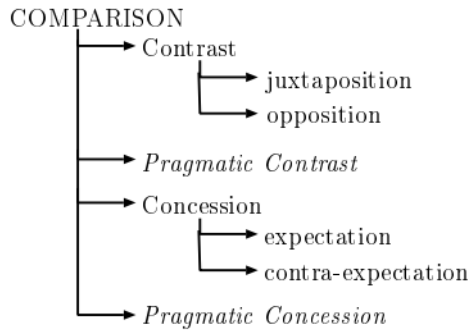


Fig. 1. A fragment of the Scheme of the sense tags for discourse relations annotation in PDTB

2.2. Studies on Contrast and Comparison relations in Russian

The majority of works on Contrast in Russian concerns a special subclass of the conjunctions used in coordination of two clauses. These are the so-called adversative conjunctions such as a ‘while, and, but’, no ‘but’ etc. (for the list of the conjunctions see, e.g. [Shvedova 1980]; [Apresyan, Pekelis 2012], [Sannikov 1989]). These conjunctions are highly polysemous (for other languages too, cf. [Spencer, Lobanova 2009]). Therefore, some other supporting linguistic features are discussed in the literature. [Apresyan, Pekelis 2012] mention discourse particles, e.g. particle *zhe* translated as ‘just, but, and’ in some dictionaries. [Shvedova 1980] and others mention grammatical parallelism and lexical oppositions. Besides, there are different additional lexical markers (referred to as ‘discourse specifiers’ further). Below we give a list of different features mentioned in the literature (Section 2.3).

Many studies are devoted to the fine-grained taxonomy of various senses of the conjunctions expressing Contrast [Apresyan, Pekelis 2012]; [Zaliznyak, Mikaelyan 2005]. There are also some general works concerning the semantics of particular conjunctions [Uryson 2004]; [Uryson 2012]; [Kobozeva 2011]; [Inkova-Manzotti 2001].

There have been attempts to formulate the notions of Contrast or Comparison in abstract terms. Thus, O. Inkova-Manzotti defines the relation of Contrast as a semantic configuration in which the Speaker represents two states of affairs as incompatible [ibid: 80]. In [Uryson, 2004] a mental operation of Comparison is described: it involves a human subject and at least two objects X and Y that have a common parameter Z (often implicit). Within this approach Contrast can also be treated a subtype of Comparison. In our discussion, we will use the notions of objects and parameters of comparison for both of the relations.

2.3. Summary of the Features Associated with Contrast

The primary cues for Contrast and Comparison are some of the coordinatives and other types of connectives, and some types of parentheticals. The following features are usually regarded as additional cues:

1. “**syntactic parallelism**” (syntactical similarity) of the coordinate components (DU), e.g. [Kreydlin, Paducheva 1974]; [Uryson 2004]; [Asher, Lascarides 2003];
2. the so called **discourse specifiers**—lexical items that are not generally recognized as DM but which tend to contribute to Contrast, e.g. *even*, *still*, *too* for English [Spenader, Lobanova 2009] and *zato* ‘but then’, *vse zhe* ‘yet’, *vdrug* ‘suddenly’, *dazhe* ‘even’ etc. [Shvedova 1980]. In RusGram [Pekelis 2018], many of these items belong to a special part of speech—correlates. [Siyjuan’, Sheremet’eva 2018] distinguish discourse specifiers used for:
3. **lexical expression of negation**: particle *ne* ‘not’ [Syjuan’, Sheremet’eva 2018], negative pronouns and some others;
4. **lexical parallelism**: synonyms and word repetitions (e.g. [Inkova-Manzotti 2001]); **antonyms, or lexical opposition**; there is some theoretical description of antonyms in Contrast, e.g. 3d type of constructions in [Siyuan’, Sheremet’eva 2018], as well as some applications of lexical oppositions to automatic Contrast detecting [Harabagiu et al. 2006]; [Marcu, Echihabi 2002]; [Murphy et al. 2009]. On the other hand, [Spenader, Stulp 2007], [Feltracco et al. 2018] note that lexical opposition is not common in cases of Contrast.

3. Data

3.1. The corpus

The current study is based on the Ru-RSTreebank (<https://linghub.ru/ru-rstreebank/>) [Pisarevskaya et al., 2017]; [Toldova et al., 2018]. It consists of 179 texts (203,287 tokens in total) and represents the genres of news and popular science (79 texts) and scientific papers (100 texts). The paper uses the most recent unpublished version of the corpus, available on request. The corpus was annotated by several annotators, with the last Krippendorff’s unitized alpha measurement of 81%, which is a good inter-annotator agreement level.

3.2. The data

For our research we use examples of Contrast and Comparison from Ru-RSTreebank. There are about 570 examples of Contrast and 234 examples of Comparison. The following manual was used in the annotation of discourse relations: https://rstreebank.ru/assets/docs/Manual_for_ru_RSTreebank_Annotation.pdf. Although in the manual three DC examples for Contrast are suggested (*a* ‘but, and’, *no* ‘but’, *nesmotrya na* ‘despite of’), DC of this relation are much more diverse (about 50 for Contrast).

3.3. Annotation for discourse relation signals

All the examples in our dataset were manually annotated for different types of linguistic signals of rhetorical relations. Firstly, the primary markers were singled out. By primary we mean those expressions that serve as overt markers for the relation between two DUs. Besides, the examples were annotated for supplementary signals listed in 2.3 and we marked different discourse specifiers (see 4.2). Then, we marked the presence of lexical repetitions, presence of synonyms (quasi-synonyms and semantically close expressions), hypernyms and opposite lexical expressions (e.g. antonyms, conversives, for the types of opposites see also [Feltracco et al. 2018]).

4. Contrast and Comparison connectives in Ru-RSTreebank

4.1. The general statistics for primary connectives and supplementary features

As it has been reported in the literature, there might be a lot of examples where the discourse relations are expressed implicitly without an overt DC (c.f. [Taboada, Das 2013]). In our set, 80 cases out of 569 (about 14%) have no overt primary markers for Contrast, while there are 65 examples out of 234 (about 28%) for Comparison. Thus, in news and scientific texts there is a tendency to express Contrast overtly. For Comparison some other devices are used.

Below, we summarise the data for different combinations of signals in Contrast vs. Comparison. We treat the syntactic parallelism separately (it was manually annotated by experts). We consider the presence of different semantically related expressions or lexical repetitions in two parts of a relation as lexical parallelism. The general statistics is given in Fig. 2 ((a) for Contrast and (b) for Comparison):

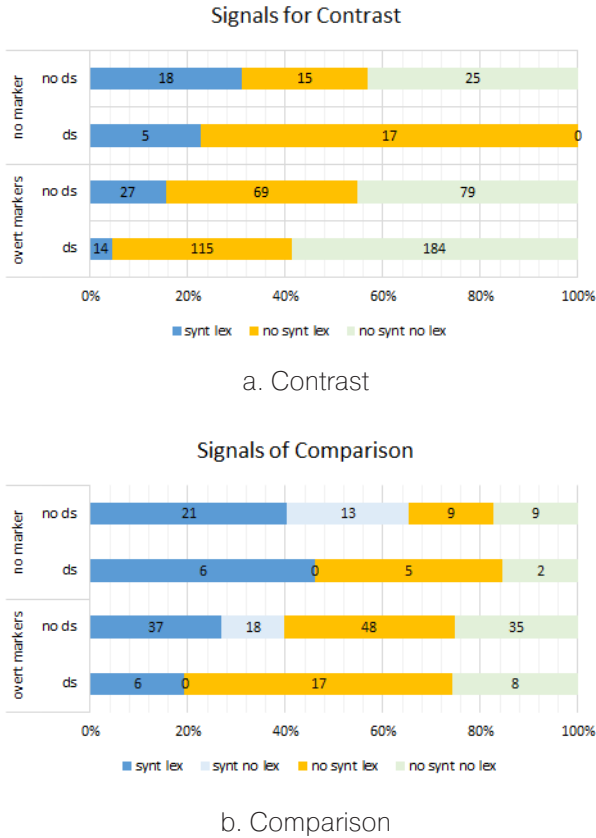
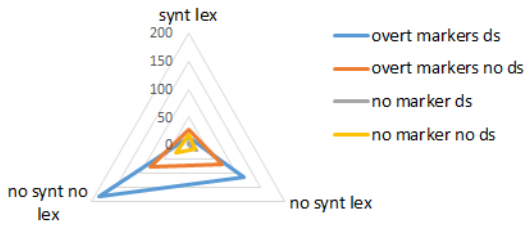


Fig. 2. The distribution of different signals for Contrast and Comparison

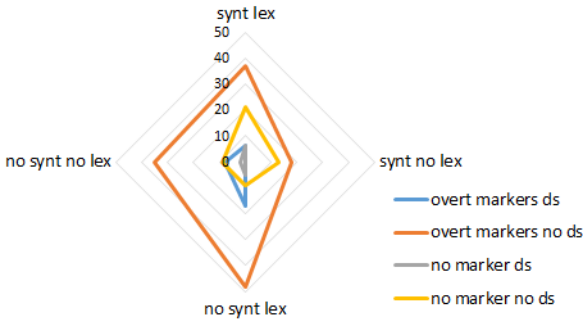
As our data shows, syntactic parallelism is more frequent with Contrast (11%); it is also quite frequent with Comparison (26%). The main tendencies can also be seen in Fig. 3. Contrast prefers the overt markers without syntactic or lexical parallelism (38%). It is supported by a large number of discourse specifiers (cf. the bleu contour corresponding to the overt markers + discourse specifier occupies the largest area and stretches towards ‘no syntactic or lexical parallelism’ direction). Moreover, lexical parallelism coincides with syntactic in Contrast. On the contrary, Comparison prefers lexical parallelism, it is quite frequent irrespective of the syntactic one, c.f. the orange radar for overt markers without discourse specifiers. The latter occupies the largest area on the chart for Comparison. The discourse specifiers are rare in Comparison (the blue radar occupies one of the smallest areas).

Main tendencies for interaction of signals for Contrast



a. Contrast

Main tendencies for interaction of signals for Comparison



b. Comparison

Fig. 3. The main tendencies in signals for Contrast vs. Comparison distribution (a radar chart)

As for primary DMs, **Table 1** provides the statistics on the most frequent DC for each relation as compared to the frequency for the opposite relation:

Table 1. DC frequency in Ru-RSTreebank

Discourse Connective	Number in Contrast	Number in Comparison
<i>no</i> 'but'	143 (28.98%)	2 (0.93%)
<i>odnako</i> 'however'	133 (27.14%)	0
<i>a</i> 'but, and'	49 (10%)	18 (13.02%)
<i>ne... a</i> 'not ... but'	30 (6.12%)	0
<i>zhe</i> "just, but, and"	13 (2.45%)	13 (6.05%)
<i>tem ne meneye</i> 'nevertheless'	12 (2.45%)	0
<i>yesli ... to</i> 'if ... then'	10 (2.04%)	4 (0.94%)
<i>v to vremya kak</i> 'while'	9 (1.84%)	9 (3.72%)
<i>(i) v to zhe vremya</i> '(and) at the same time'	6 (1.22%)	3 (1.41%)

Discourse Connective	Number in Contrast	Number in Comparison
<i>s drugoy storony</i> ‘on the other hand’	6 (1.22%)	1 (0.47%)
<i>khotya</i> ‘although’	6 (1.22%)	0
<i>chem ... tem</i> ‘the ... the’	0	6 (2.79%)
<i>v otlichije ot X</i> ‘unlike X’	4 (0.82%)	6 (2.79%)
<i>analogichnyy</i> ‘similar’	0	4 (0.94%)
<i>kak ... tak i</i> ‘as ... so’	0	4 (0.94%)
<i>i</i> ‘and’	1 (0.20%)	4 (0.94%)

Other connectives for Contrast occur in less than 1% of cases, e.g. *v protivnom sluchaye* ‘otherwise’, *naprotiv* ‘on the contrary’, *togda kak* ‘whereas’, *v otlichije ot X* ‘in contrast to’, *vmesto X* ‘instead X’, *nesmotrya na X* ‘despite of X’, *vprochem* ‘however’ etc.

There are specific DMs for Comparison in our data set that express the similarity of two entities’ properties or two states of affairs like *analogichnyj* ‘analogous’, *pokhozhiy* ‘similar to’, *takoj zhe* ‘similar’, *kak i* ‘as’. Correlative constructions (DCs consisting of two parts situated in different text spans) are among these markers (e.g. *s odnoy storony ... (a) s drugoy (storony)* ‘on the one hand ... (a) on the other (hand)’). The majority of correlatives found in our data serve as DMs for both relations.

While Contrast has quite a distinct DM profile (the two most frequent DMs cover more than 50% examples), there is no dominant specific DM for Comparison. The most frequent conjunction covers only 13% of examples. Moreover, it is used for Contrast in the same proportion of cases. The DM diversity for Contrast is 0,1 (48 markers for 464 cases with overt DM), and it is higher for Comparison, its value is 0,5 (90 markers for 169 cases).

Another important Contrast property is that it is often marked by a sequence of connectives. One of them is a simple conjunction, the second one can be a multiword parenthetical expression or a series of particles (e.g. *no* ‘but’ + *tem ne meneye* ‘nevertheless’):

Although there are specific DCs for each relation, there are a lot of connectives in our data that occur both with Comparison and Contrast, which means that there should be other signals that help to differentiate the two relations.

4.2. Lexical specifiers

[Shvedova 1980] and others (see also 2.2) report the usage of additional expressions specifying some aspects of the opposed states of affairs. These discourse specifiers can be helpful for differentiating the two relations. The general statistics in 4.1 shows that Contrast usually has these additional expressions (about 85% of cases for Contrast, while only 7% for Comparison). Below are some examples of discourse specifiers, grouped in different classes:

- a) **temporal words** and constructions—*v to zhe vremya* ‘at the same time’, *vse yeshche* ‘still’, *uzhe* ‘already’; their function is to mark the unexpected simultaneous existence of two opposed states of affairs or to emphasize the opposition of the states of affairs in different time slots.

- b) markers of **epistemic modality**—*deystvitel'no* ‘really’, *bezuslovno* ‘absolutely’, *yavnyy/yavno* ‘explicit’; they emphasize the reality of an unexpected state of affairs;
- c) **anaphoric expressions, e.g. different types of pronouns and quantifiers**: *etot* ‘this’, *dannyi* ‘given’, *podobnyy* ‘similar’, *drygoj* ‘other one’; one of their functions is to mark the NPs whose properties are contrasted (e.g. contrastive topics); there are many examples where they go in pairs (c.f. *odin/etot*—*drugoj* ‘one of them/this—the other one’, *nekotoryje* ‘some of them’—*drugije* ‘others’)
- d) content words such as **verbs of contradiction**—*vozzrazhat'* ‘to object’, *protivorechit'* ‘to contradict’;
- e) focus and topic re-activation **particles**—*imenno* ‘precisely’, *zhe* ‘but, as to’, *dazhe* ‘even’, *lish'* ‘only’, *tol'ko* ‘only’, *vse ravno* ‘still’.

A special attention should be drawn to the proportion of negative particles **ne** ‘not’ and other words of negation (e.g. *nevozmozhno* ‘impossible’), negative pronouns etc.

4.3. Lexical parallelism

In accordance with the definitions, relations Comparison and Contrast involve objects (X and Y) that are compared, and a parameter of comparison (Z). The implication is that some of the lexemes in the two parts of the relations are semantically related. Lexical repetitions and synonyms can specify the general grounds of comparison and the opposite notions express the difference. We treat synonymy and semantic opposition very widely here (see [Lyons 1977]). **Table 2** shows the statistics on different types of lexical parallelism in our data:

Table 2. The distribution of different types of lexical parallelism in Contrast vs. Comparison

	Contrast	Comparison
Semantically opposed expressions	200 (35%)	42 (18%)
(Quasi)-synonyms, hypernyms	63 (17%)	94 (36%)
Repeated and cognate words	142 (25%)	62 (23%)
Total	280	108

Synonyms and repetitions are more typical for Comparison cases against (X-squared = 43.595, df = 1, p-value = 4.039e-11 (Pearson’s Chi-squared test with Yates’ continuity correction). They can be both the objects (X/Y) and the parameter of comparison (Z) (4), while in Contrast they are used for denoting only the objects of comparison (5):

- (4) [Naiboleye kharakternymi chertami obraza ideal'noy zhenshchiny *russkoyazychnogo reklamnogo parfyumernogo diskursa* (X) *yavlyayutsya: privilekatel'nost', seksual'nost', zhenstvennost'* <...> (Za).] [Bol'shinstvo *vysheperechislennykh priznakov kharakterny i dlya nemetskoyazychnogo diskursa* (Y). *Predstavleniya ob ideal'noy zhenshchine takzhe svyazany s takimi ponyatiyami, kak privilekatel'nost', <...> unikal'nost'* (Zb)]

‘The most characteristic features of the image of an ideal woman of the Russian-language advertising perfume discourse (X) are: attractiveness, sexuality, femininity <...>. (Za) Most of the above symptoms are characteristic of the German-language discourse (Y). Ideas about the ideal woman are also associated with such concepts as attractiveness, <...> uniqueness. (Zb)’

- (5) [*Vo frantsuzskom yazyke* (X) *prevalliruyut zvuki, obrazuyemyye v peredney chasti golosovogo apparata* (Za).] [*V angliyskom yazyke* (Y), *naprotiv, preobladayut glasnyye zvuki zadnego ryada* (Zb).]
‘In French (X), the sounds formed in front of the voice apparatus (Za) prevail. In English (Y), on the contrary, backward vowel sounds (Zb) prevail.’

On the contrary, semantically opposed expressions can denote both X and Y or Za and Zb for Contrast, while they stand for the object of comparison only in Comparison.

Besides, Contrast can be established between the two DUs where the opposition is not between the two states of affairs, but between the implicit expectations of the first DU and the real state of affairs (stated in the second DU). For these cases no lexical parallelism is expected:

- (6) [*Privivki ostayutsya naiboleye effektivnoy meroy preduprezhdeniya epidemiologicheskikh zabolevaniy.*] [*Odnako mnogiye roditeli otkazyvayutsya vaktsinirovat’ svoikh detey.*]
‘Vaccinations are still the most effective preventive measure for epidemiological diseases.’ ‘However, many parents refuse to vaccinate their children.’

In (6) the second span is in opposition not with the first one as such, but with its implication that has no explicit expression—‘people need to be vaccinated to prevent diseases’. As the result, there are fewer repetitions of words in Contrast, since the second segment does not contrast directly with the first one.

5. Conclusion

In this paper, we present an analysis of different signals of the Contrast relation in Russian as compared to those used for Comparison. We examine different cues mentioned in the literature as marking the corresponding relations against our corpus data. For this purpose, we have annotated different types of signals for all the examples. As a result, we compiled a list of Contrast and Comparison DMs based on RuRSTreebank (48 and about 90 elements respectively).

The two relations under discussion have much in common, e.g. the DC used in both relations are often ambiguous or they can be expressed implicitly. In these cases, subsidiary signals are used. These are special lexical markers (particles, adverbs etc.), syntactic and lexical parallelism. Because of these similarities, it may be difficult to draw a line between the two relations. The distinction between these two relations should be presented as a scale rather than a mere dichotomy.

However, our data confirm the role of various additional signals for detecting Contrast and differentiate it from Comparison, like discourse specifiers, markers of negation and semantically opposed expressions.

References

1. *Apresyan V. Yu., Pekelis O. Ye.* (2012), Coordinate conjunctions [Sochinitel'nyye soyuzy]. Materials for the project of Russian grammar corpus description [Materialy dlya proyekta korpusnogo opisaniya russkoy grammatiki], available at: <http://rusgram.ru>, M., As a manuscript [Na pravakh rukopisi].
2. *Asher N., Lascarides A.* (2003), Logics of conversation. Cambridge University Press.
3. *Carlson L., Marcu D.* (2001), Discourse tagging reference manual. Technical report, Information Science Institute, Marina del Rey, CA. ISI-TR-545.
4. *Feltracco A., Magnini B., Jezek E.* (2018), Lexical Opposition in Discourse Contrast, Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Torino, Italy, December 10–11.
5. *Harabagiu S., Hickl A., Lacatusu F.* (2006, July), Negation, contrast and contradiction in text processing, AAAI, Vol. 6, pp. 755–762.
6. *Hobbs J. R.* (1985), On the coherence and structure of discourse. Technical Report 85–37, Center for the Study of Language and Information (CSLI), Stanford, CA.
7. *Inkova-Manzotti O.* (2001), Opposition connectives in French and Russian [= Connecteurs d'opposition en français et en russe] [Konnektory protivopostavleniya vo frantsuzskom i russkom yazykakh], M., Informelektro.
8. *Kobozeva I. M.* (2011), Conjunctions as markers of rhetorical relations in discourse: Russian conjunction “i” [Soyuzy kak markery ritoricheskikh otnosheniy v diskurse: russkiy soyuz «i»], *L'analisi linguistica e letteraria*, 19, № 2, pp. 365–387.
9. *Kreydlin G. E., Paducheva E. V.* (1974), Semantics and syntactic properties of conjunction ‘a’ [Znachenije i sintaksicheskiye svoystva soyuza a]. Scientific and technical information [Nauchno-tehnicheskaya informatsiya], Seriya 2, 9, pp. 31–37.
10. *Laali M., Kosseim L.* (2017), Automatic Mapping of French Discourse Connectives to PDTB Discourse Relations, Proceedings of the SIGDIAL 2017 Conference, Saarbruecken, Germany, pp. 1–6.
11. *Lyons J.* (1977), Semantics. V. 1. Cambridge University Press.
12. *Mann W. C., Thompson S. A.* (1988), Rhetorical Structure Theory: Toward a Functional Theory of Text Organization, Text 8, 3, pp. 243–281.
13. *Marcu D., Echihabi A.* (2002), An unsupervised approach to recognizing discourse relations, Proceedings of the 40th annual meeting of the association for computational linguistics.
14. *Mendes A., del Rio I., Stede M., Dombek F.* (2018), A Lexicon of Discourse Markers for Portuguese—LDM-PT, Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018), Miyazaki, Japan, pp. 4379–4384.
15. *Murphy M. L., Paradis C., Willners C., Jones S.* (2009), Discourse functions of anonymy: A cross-linguistic investigation of Swedish and English, Journal of pragmatics, 41(11), pp. 2159–2184.
16. *Pekelis O. Ye.* (2018), Correlates (that, so and others) [Korrelyaty (to, tak i dr.)] Materials for the project of Russian grammar corpus description [Materialy dlya proyekta korpusnogo opisaniya russkoy grammatiki], available at: <http://rusgram.ru>. As a manuscript [Na pravakh rukopisi], M, 2018.

17. *Pisarevskaya D., Ananyeva M., Kobozeva M., Nasedkin A., Nikiforova S., Pavlova I., Shelepov A.* (2017), Towards building a discourse-annotated corpus of Russian, *Computational Linguistics and Intellectual Technologies. Proceedings of the International Conference "Dialogue 2017"*, pp. 194–204.
18. *Prasad R., Miltsakaki E., Dinesh N., Lee A., Joshi A., Robaldo L., Webber, B. L.* (2007), *The penn discourse treebank 2.0 annotation manual.*
19. *Sannikov V. Z.* (1989), Russian coordinate constructions: semantics, pragmatics, syntax [Russkiye sochnitel'nyye konstruksii: semantika, pragmatika, sintaksis], M, "Nauka".
20. *Shvedova, N. Yu. ed.* (1980), *Russian grammar [Russkaya grammatika]. V dvukh tomakh.* AN SSSR Institut russkogo jazyka, M.: Nauka, 1980.
21. *Siyuan' K. H., Sheremet'yeva Ye. S.* (2018), The specificity of structures built on the basis of a combination of union and with the concretizations 'in fact', 'essentially', 'in essence' [Spetsifika konstruksiy, stroyashchikhsya na osnove sochetaniya soyuza s konkretizatorami posuti, posushchestvu, vsushchnosti], *Scientific dialogue [Nauchnyy dialog]* (1).
22. *Spender J., Lobanova A.* (2009, January), Reliable discourse markers for contrast relations. In *Proceedings of the Eighth International Conference on Computational Semantics* (pp. 210–221), Association for Computational Linguistics.
23. *Spender J., Stulp G.* (2007, January), Antonymy in contrast relations, *Seventh International Workshop on Computational Semantics*, Vol. 3, p. 100.
24. *Taboada M., Das D.* (2013), Annotation upon annotation: Adding signalling information to a corpus of discourse relations, *Dialogue and Discourse* 4(2), pp. 249–281.
25. *Taboada M., Mann W. C.* (2006), Rhetorical structure theory: Looking back and moving ahead, *Discourse studies*, 8(3), pp. 423–459.
26. *Toldova S., Kobozeva M., Pisarevskaya D.* (2018), Automatic mining of discourse connectives for Russian, *Conference on Artificial Intelligence and Natural Language*, pp. 79–87.
27. *Uryson E. V.* (2004), Some meanings of conjunction A in the light of modern semantic theory [Nekotoryye znacheniya soyuza A v svete sovremennoy semanticheskoy teorii]. *Russian language in scientific coverage [Russkiy yazyk v nauchnom osveshchenii]*, (2), 17.
28. *Uryson E. V.* (2012), Conjunctions, connectives, and the valence theory [Soyuzy, konnemy i teoriya valentnostey]. In *Computational linguistics and intellectual technologies [Komp'yuternaya lingvistika i intellektual'nyye tekhnologii]*, pp. 627–638.
29. *Webber B. L., Joshi A. K.* (1998), Anchoring a lexicalized tree-adjointing grammar for discourse. *Discourse Relations and Discourse Markers: Proceedings of the Conference*, Association for Computational Linguistics, pp. 86–92.
30. *Zaliznyak, A. A., Mikaelyan, I. L.* (2005) Russian Conjunction "a" as a language-specific word [Russkiy soyuz a kak lingvospetsifichnoye slovo], *Computational Linguistics and Intellectual Technologies, Proceedings of the International Conference "Dialogue 2005"*, pp. 153–159.