

Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2018”

Moscow, May 30—June 2, 2018

THE CUES FOR RHETORICAL RELATIONS IN RUSSIAN: “CAUSE—EFFECT” RELATION IN RUSSIAN RHETORICAL STRUCTURE TREEBANK¹

Toldova S. (toldova@yandex.ru)

NRU Higher School of Economics, Moscow, Russia

Pisarevskaya D. (dinabpr@gmail.com),

Kobozeva M. (kobozeva@isa.ru)

Institute for Systems Analysis FRC CSC RAS, Moscow, Russia

Vasilyeva M. (linellea@yandex.ru)

Lomonosov Moscow State University, Moscow, Russia

The purpose of the paper is to investigate cues signalling the relations between discourse units in Russian. Building a lexicon of discourse connectives is an indispensable subtask in many discourse parsing applications as well as an essential issue in theoretical researches of text coherence. In order to develop such a resource for Russian, we have conducted a corpus-based study of discourse connectives that were manually extracted from the Russian Rhetorical Structure Treebank (Ru-RSTreebank). The Treebank includes 79 texts annotated within the RST framework (Mann, Thompson 1988). In order to provide a deeper analysis of connectives in Russian, we focus on causal relations only, namely, the ‘Cause-Effect’ relation. Some of the connectives (primary connectives) are enumerated in grammars and dictionaries. They primarily mark the intra-sentential relations. However, there is an expansive class of less grammaticalized items (secondary connectives) that have received less attention till now. Some of them are based on content words (e.g. по причине ‘for the cause’). Secondary connectives often serve as linking devices for inter-sentential relations.

We suggest a scheme for connectives annotation for Russian. We specify the basic patterns that can be used for less-grammaticalized connectives mining in an unannotated corpus. Besides, we provide the comparison of two classes of connectives (primary vs. secondary ones). Our research

¹ The study was funded by RFBR according to the research project № 17-29-07033

has shown that these two classes differ in their properties. There is a statistically significant difference between them with respect to the nucleus/satellite position, intra- vs. inter-sentential relations and some others.

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

МАРКЕРЫ РИТОРИЧЕСКОГО ОТНОШЕНИЯ «ПРИЧИНА — СЛЕДСТВИЕ» В РУССКОМ ЯЗЫКЕ НА МАТЕРИАЛЕ КОРПУСА РИТОРИЧЕСКИХ СТРУКТУР

Толдова С. (toldova@yandex.ru)

НИУ ВШЭ, Москва, Россия

Писаревская Д. (dinabpr@gmail.com),

Кобозева М. (kobozeva@isa.ru)

Институт системного анализа ФИЦ ИУ РАН,

Москва, Россия

Васильева М. (linellea@yandex.ru)

МГУ имени М. В. Ломоносова, Москва, Россия

Создание лексикона дискурсивных коннекторов является одной из актуальных задач при разработке систем автоматического анализа дискурса. Описание коннекторов также играет немаловажную роль в теоретических исследованиях связности текста. В целях создания соответствующего лексикона для русского языка мы провели корпусное исследование коннекторов, выделенных экспертами в корпусе Ru-RSTReebank. Этот корпус представляет собой 79 научно-популярных и новостных текстов, размеченных в терминах теории риторических структур (Mann, Thompson 1988). Вопрос о том, как устроен класс маркеров риторических отношений в русском рассматривается на примере каузальных отношений, в частности, на примере отношения “причина-эффект”. Некоторые коннекторы (первичные коннекторы) представлены в грамматиках и словарях. Как правило, они маркируют связи внутри предложения. Однако существует достаточно обширный класс менее грамматикализованных коннекторов (вторичные коннекторы), которые исследованы в меньшей степени. В частности, в качестве коннекторов используются конструкции с полнозначными лексическими единицами (например, по причине). Многие из таких коннекторов маркируют связи между предложениями и дискурсивными единицами большего объема. Таким образом, настоящая работа

посвящена анализу коннекторов, которые обеспечивают связь между дискурсивными единицами в русском языке. Особое внимание уделяется менее грамматикализованным коннекторам, в том числе коннекторам, обеспечивающим связность на меж-сентенциальном уровне. В работе мы предлагаем схему описания маркеров риторических отношений, разработанную на основе проведенного анализа, описываем основные модели образования свободных конструкций, с помощью которых список коннекторов может быть расширен с использованием неразмеченного корпуса текстов. Также в статье рассматриваются результаты сравнения двух классов коннекторов (первичных и вторичных). Между данными классами наблюдается статистически значимая разница в отношении ряда признаков, таких, например, как положение внутри ядра/сателлита, тенденция к маркированию внутрисентенциальных vs. меж-сентенциальных отношений и др.

Ключевые слова: дискурсивный анализ, теория риторических структур, дискурсивные коннекторы, корпусная лингвистика, корпусная разметка

1. Introduction

The analysis of discourse structure is a challenging issue for linguistic theory. It plays an important role in many high-level NLP applications, such as text summarization [Louis et al. 2010], sentiment analysis [Voll and Taboada 2008], question answering [Ferrucci et al. 2010], argumentative discourse analysis [Galitskij et al. 2018] and others. Discourse parsing presupposes establishing the relations between coherent text spans. In many approaches, the identification of these relations relies on detecting special lexical clues. Thus, constructing a lexicon of discourse connectives is an essential task.

In this paper, we present a corpus study of discourse connectives for Russian from the perspective of constructing such a lexicon. As a source of data, we use the pilot Russian RST Treebank, built in 2017². To provide an in-depth analysis of connectives, we focus on causal relations with a special emphasis on 'Cause-Effect' relation.

We consider two basic classes of connectives, namely, primary vs. secondary ones depending on whether they are registered in Russian grammars or not (cf. primary vs. secondary connectives distinction in [Rysová M., Rysová K. 2014]).

The first class includes among others different functional words such as conjunctions, prepositions etc. They have been studied for many years [Shvedova ed. 1980; Pekelis 2014, etc.]. However, the role of these connectives as signals of rhetorical relations within the RST framework still remains under-investigated. Besides, little attention was paid in the literature to less grammaticalized items that can serve as clues for inter-sentential relations.

As the result of our study, we provide the analysis of connectives with a special emphasis to less-grammaticalized items. We suggest a scheme for connectives annotation for Russian based on our corpus study with due consideration of other approaches

² <http://linghub.ru/ru-rstreebank/>

[Roze et al. 2012]; [Stede, Umbach 1998]; [Mírovský et al. 2017] etc. We specify the basic patterns for less-grammaticalized connectives. These patterns can be used for extracting new connectives from an unannotated corpus. Besides, we provide a comparison of two classes of connectives (primary vs. secondary ones). Our research has shown that these two classes differ in their properties.

2. Background

In our research, the underlying **discourse structure representation** is the *Rhetorical Structure Theory* [RST; Mann, Thompson 1988]. It assumes, that a text is organized into a hierarchical non-projective tree where discourse units (text spans) of smaller size are embedded into bigger ones. Discourse units are connected to each other by rhetorical relations. We concentrate our attention only on asymmetric relations, in which one of the text spans, the Nucleus, carries more important information than the other one, the Satellite, as in *(Peter went home)_{nucleus} (because he was tired)_{satellite}* [Mann, Thompson 1988].

As we are dealing with written texts, we consider clauses as elementary discourse units (EDUs), and not prosodic units (as in [Hirschberg, Litman 1993]; [Chafe 1994]; [Kibrik, Podlesskaya 2003]). Structures smaller than a finite clause, such as nominalized constructions or infinitival clauses, can also be treated as EDUs [Carlson, Marcu 2001]; [Schauer 2000]. For example, a preposition can signal causal relations between its dependant expressed via nominalization and the rest part of a clause as in *Из-за (его позднего возвращения)...* ‘due to his late return’).

The rhetorical relations quite often are signalled via special lexical clues (discourse connectives). Thus, the **construction of a discourse connectives lexicon** is an essential task. One of the possible approaches is to compile a repository of connectives manually, using standard dictionaries and grammars (e.g. dictionary of connectives for German and English - DiMLex [Stede, Umbach 1998]), for Spanish [Alonso et al. 2002], for French [Roze et al. 2012]. Another way to compile a list of connectives is extracting them from available corpora, e.g. Arabic lexicon [Al-Saif et al. 2010], the list of connectives for Russian [Toldova et al. 2017]. Finally, a list of connectives extracted from existing corpora for a source language can be translated into the target language [Meyer, Webber 2013].

Another relevant task is to settle a set of **annotation features** for **connectives classification**. In [Grote, Stede 1998] authors propose **syntactic** features such as part-of-speech, type of connection it establishes, scope of a connective, linear ordering of the conjuncts, connective position within a text span, **semantic** (semantic relations, polarity, functional ordering of spans) and **pragmatic** ones (discourse relation, presupposition, stylistic features). Some contextual features (occurrence in initial/final sentence or segment, previous/following word, level of embedding etc.) are mentioned in [Alonso et al. 2002]. In the Penn Discourse Treebank (PDTB) approach (cf. international multilingual project TED [Lee et al 2016]), discourse connectives are treated as discourse-level predicates that have two arguments – text spans referring to events or states [Prasad et al. 2007]. We consider this approach while working out our own scheme of connectives annotation.

We also take into account the typology of causal relations signals [Asghar 2016; Chang, Choi 2006]; [Khoo 1998]. According to [Khoo 1998], these are the following types of devices: (1) causal connectives linking two phrases, clauses or sentences (adverbial, prepositional, clause-integrated connectives); (2) causative verbs - transitive verbs that specify the result of an action, event or state, or the influence of some object; (3) resultative constructions; (4) conditional constructions; (5) causative adverbs and adjectives. Only type (1) and type (2) devices are represented in our corpus in sufficient quantity.

Thus, connectives lexicons usually represent morphological, syntactic, semantic and pragmatic description of connectives, the constraints on linear position of discourse units they connect and some other configurational properties.

3. Data

3.1. The current study is based on the **relations annotated in the Ru-RSTreebank** [Pisarevskaya et al. 2017]. The corpus consists of 79 texts, including news, news analytics and popular science (5582 EDUs and 49,840 tokens in total). The text segmentation in the corpus satisfies the principles suggested in [Carlson, Marcu 2001]. Besides EDUs, corresponding to finite clauses, it contains intra-clausal EDUs [cf. Schauer, 2000]. Thus, prepositional phrases, adverbial phrases headed by corresponding connectives (cf. *because of*, *in spite of*) are treated as separate EDUs.

We limit our investigation only to one type of discourse relations, namely, causal ones. There are 220 examples for the Cause-Effect relation and 110 for the Evidence in our corpus, including both intra- and inter-sentential relations. All these examples are annotated in terms of discourse connectives and their properties.

3.2. The list of **connectives** is manually extracted from the examples. The connectives are divided into two classes, primary vs. secondary connectives, according to their degree of grammaticalization [cf. Rysová and Rysová, 2014, 2015]. As the degree of grammaticalization is a gradable feature rather than a binary one, we rely on the Russian grammar [Shvedova ed. 1980] as a reference source.

Primary connectives are “mainly grammatical (or functional) words which primary function is to connect two units of a text” [Rysová M., Rysová K. 2014], e.g. *из-за* ‘because of’, *поэтому* ‘therefore’ etc. We treat the connectives that are enumerated in Russian grammar [Shvedova ed. 1980] as primary connectives. Thus, some of the multi-word prepositions are also treated as primary connective (e.g. *в связи с* ‘in connection with’).

Secondary connectives, known as Alternative Lexicalization (AltLex) in the Penn Discourse Treebank [Prasad et al. 2010], are not yet fully grammaticalized. These are, primarily, multi-word expressions, e.g. *это привело к тому, что* ‘this led to the fact that’, *причина этого...* ‘the cause is...’ etc. These connectives are quite frequent in our corpus. They occur in 46.5% of our examples.

Thus, the class of secondary connectives is of special interest. It constitutes a heterogeneous and open-ended class of elements. Our goal is to single out basic patterns for secondary connectives formation.

The final list of discourse connectives for causal relations consists of 48 connectives (see Fig. 1 for the most frequent of them).

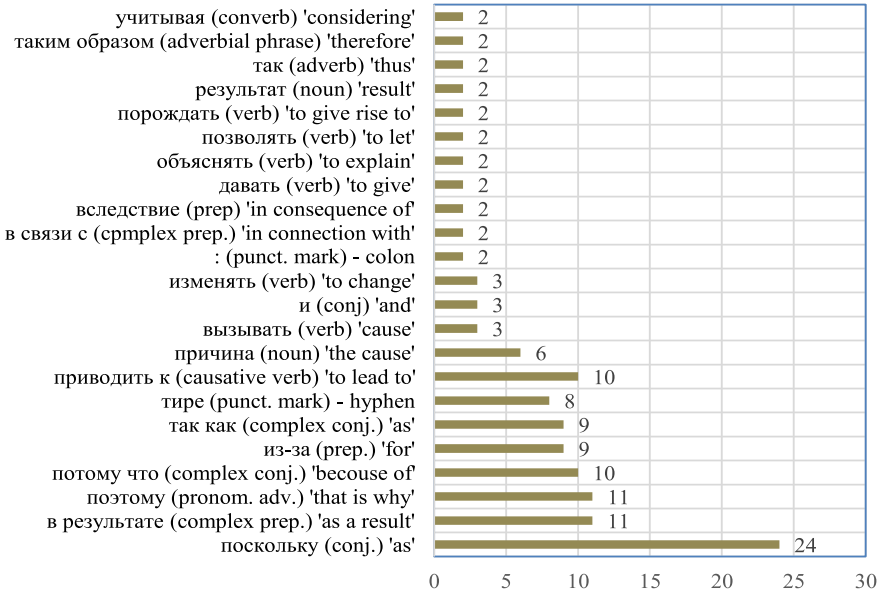


Figure 1. Discourse connectives frequency in Ru-RSTreebank (frequency > 1)

3.3. Relying on the essential works devoted to the construction of connective lexicons [Grote, Stede, 1998]; [Alonso et al. 2002]; [Mírovský et al. 2017], we suggest a scheme for connectives annotation and annotate all the occurrences of connectives in our examples. The scheme includes the properties of connectives (whether it is a multi-word expression (MWU) or not, POS of its core word (for core word see 4.1.); (b) whether a connective is mentioned in the certain resources (see 3.4 for details); (c) the position of a connective in a clause and in a sentence; (d) the properties of the arguments of a connective, such as their complexity, their position and their grammatical features (whether they are headed by a finite verb or by non-finite verb forms). Our final set of features is presented in Table 1.

Table 1. The annotation scheme with an example: the annotation of 'Что приводит к тому, что' [that leads to...]

Feature	Values	Example
Type of connective	primary/secondary/NA	secondary
Simple or complex structure	simple/compound	MWU
Listed in the RNC MWU lists	yes/no	no
A causal conjunction listed in RusGram	yes/no	no

Feature	Values	Example
Core word in a connective	verb	verb
Position of a connective within EDU	clause initial / clause internal	clause initial
Position of a connective within a sentence	sentence initial/no	no
Position of a connective wrt. nucleus vs. satellite	nucleus/satellite	nucleus
Connection type wrt sentential boundaries	intra-/inter-sentential	inter-sentential
Order of Nucleus (wrt. satellite)	1/2	
Occurrence with an anaphoric element	anaphoric element	yes
Satellite complexity	span, sentence, multiclausal, clause	clause
Nucleus complexity	span, sentence, multiclausal, clause, subclause	clause
Number of clauses in the Nucleus	number	1
Number of clauses in the Satellite	number	1
Gram. features of the Nucleus head	Indicative/converb/participle/nominalization/noun	indicative
Gram features of the Satellite head	Indicative/converb/participle/nominalization/noun	indicative
Anaphoric element in the Nucleus	this/that/what/nd	what
Anaphoric element in the Satellite	this/that/what/nd	nd
Discontinuity	no	no

We follow [Rysová M., Rysová K. 2014] in that we include the anaphoric elements into the connective annotation as in *в следствии этого* 'in consequences of this'. We take into account the demonstrative *это* 'this' (*Это могло стать причиной* 'This could have caused'), wh-relative pronouns as *что* (*Что могло стать причиной*) and expressions with *то* 'that' (cf. *то, что* 'the fact, that'). We also register what text span (Nucleus vs. Satellite) an anaphoric element refers to (see 4.2. for details).

3.4. We compare our results with the theoretical works devoted to expression of causal relations in Russian. We check the resulting list of connectives against the existing resources for functional words and phrases.

There is a detailed survey of causal subordinate conjunctions in RusGram³ [Pekelis 2014], both simple (*так как; поскольку* 'as, since' etc.) and complex

³ http://rusgram.ru/%D0%9F%D1%80%D0%B8%D1%87%D0%B8%D0%BD%D0%BD%D1%8B%D0%B5_%D0%BF%D1%80%D0%B8%D0%B4%D0%B0%D1%82%D0%BE%D1%87%D0%BD%D1%8B%D0%B5#12

(*благодаря тому* (,) *что* ‘due to’; *в результате того* (,) *что* etc. ‘as a result’). This survey concerns inter-clausal relations, though the author mentions that some of the connectives can signal intra-clausal relations, e.g. *В результате этого* ‘As a result of this’. We also have checked our list of connectives against the Russian National Corpus lists of functional MWUs⁴.

A list of content words expressing cause-effect relation is given in Ju. Apresyan [Apresyan 2001], such as causal verbs (e.g. *вызывать* (*болезнь*) ‘to cause (a disease)’, *внушать* (*ужас*) ‘to excite (a horror)’ and nouns (e.g. *основание* ‘a ground’, *мотив* ‘a motive’). The detailed analysis of lexemes expressing “cause” or “purpose” is proposed in [Boguslavskaya, Levontina 2004]. This work is devoted to lexicographic issues. However, both works provide lists of potential content words for ‘cause-effect’ connectives.

4. Causal connectives in Ru-RSTreebank and their properties

4.1. Patterns for Secondary Discourse Connectives signalling causal relations

As a result of corpus analysis, the basic patterns for secondary connectives formation were distinguished. Our classification is based on the part of speech of the core word. According to [Mírovský et al. 2017], the core word of a connective is “the word that most strongly signals the relation that the whole connective expresses”. The types of secondary connectives are presented below.

Constructions containing causative verbs; as causative verbs we treat the verbs whose meanings include a causal element [Asgar 2016] such as verbs of causation (*X позволяет Y* ‘X enables Y’, *X вызывает Y* ‘to produce’ etc.), verbs of mental impact [Paducheva 2004]; [Glovinskaya 1993] (c.f. *X можно объяснить Y-ом* ‘one can justify X via Y’), motion causation verbs (*X приводит к Y* ‘to bring about’), change of state causation verbs (*X изменяет Y* ‘X causes the change in Y’) and some other (*X порождает Y* ‘X gives rise to’):

- (1) [*Неудачно остановившаяся машина стала помехой для быстрых кругов многих гонщиков, включая Фернандо Алонсо,*] [*и это вызвало расследование FIA.*]
‘The poorly stopped car became a hindrance to the fast laps of many racers, including Fernando Alonso, and **this evoked** an investigation by the FIA.’

In (1) the connective *X вызывает Y* is in nucleus, *Y* is a nominalization while *X* is an anaphoric element *это* ‘this’, both elements *X* and *Y* are located in the same EDU (nucleus), demonstrative *это* substitutes the satellite proposition.

Light verbs constructions, that are structures including a content noun denoting ‘cause’ or ‘effect’ (*причина* ‘cause’, *результат* ‘result’, *повод* ‘matter’, *основа* ‘basis’, *основание* ‘basis’, *вывод* ‘conclusion’, *отправная точка* ‘starting point’,

⁴ <http://ruscorpora.ru/obgrams.html>

подтверждение 'confirmation', *довод* 'argument', *свидетельство* 'evidence' etc.) and a light verb (*являться* 'to be', *становиться* 'to become', *давать* 'to give' etc.):

- (2) [*... повышение цен стало результатом*] [*удорожания сырья.*]
'... the price increase **was the result** of the rise in price of raw materials.'

In (2) both arguments of the connective are nominalizations.

Complex prepositions, or secondary prepositions in term of the CzeDLex [Rysová, Rysova 2014], that usually are composed of a preposition and a content noun (*в результате X* 'as a result', *вследствие X* 'in consequence of', *в отместку за X* 'in revenge'), excluding prepositions mentioned in [Shvedova ed. 1980]:

- (3) *К тому же неприличным был объявлен текст песни «Ich tu dir weh», в результате чего композицию запретили для исполнения на публике.*
'In addition, the lyrics of the song "Ich tu dir weh" were declared indecent, **as a result of which** the composition was banned for performance in public.'

Adverbials: adverbial phrases [Kustova 2017] and converbs; (*X связан с Y* 'concerned with', *X обусловлен Y* 'caused by', *учитывая X* 'taking into account'):

- (4) *Предполагают, что синий цвет внешнего кольца обусловлен тем, что оно в дополнении к пыли обладает некоторой примесью мелких частиц водяного льда с поверхности Маба.*
'It is assumed that the blue color of the outer ring **is due to the fact that** in addition to the dust it has some admixture of small particles of water ice from the surface of the Mab.'

Adverbs ((*и*) *поэтому*+CP '(and) therefore', (*а*) *потому* '(and) that is why', *не случайно*+CP 'not accidentally'):

Other constructions with nouns (*X одна из основ Y* 'one of the bases', *как следствие* 'as a consequence', *причина ADJ*: 'a cause is ADJ'):

- (5) *Нынешний норматив — 75% — был введен законом с подачи ЦБ вскоре после «великого дефолта». Причина простая: финансовый кризис лишил его валютных резервов.*
'Current norm—75% was established by the law at the CB behest soon after the "great default". **The reason (for that) is simple...**'

The proportion of different types of connectives, with respect to the core word, is given in Fig.2:

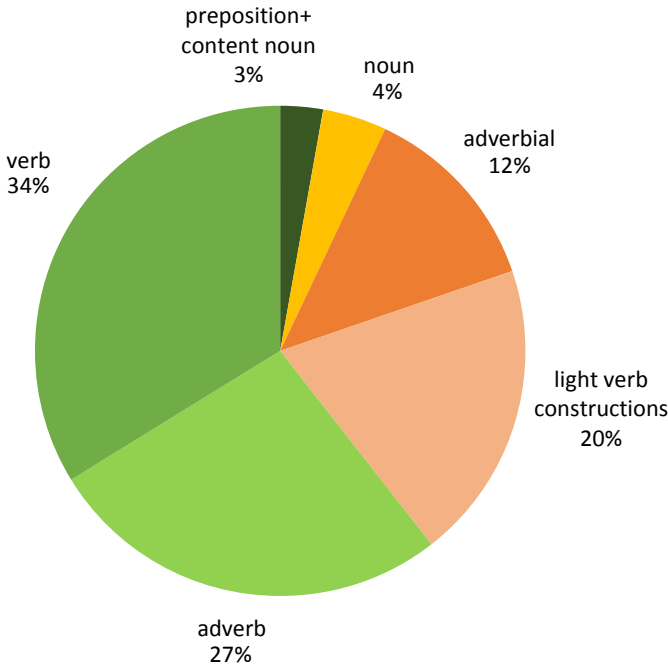


Figure 2. Core words of secondary connectives for 'Cause-Effect' relation (the grammaticalized complex prepositions 'preposition+content noun' are excluded)

4.2. Anaphoric Elements in Multi-Word Connectives

As it has been mentioned, the connectives can contain **anaphoric and cataphoric expressions** such as the demonstrative *это* (*этом*) 'this' or the relative pronoun *что* 'what':

- (6) *В то же время некоторые ведомства, в частности МЧС, готовы экспериментировать ... **Благодаря этому** они надеются резко повысить зарплату госслужащим* [НКРЯ]
 'At the same time some of the departments are ready to carryout experiments... **Due to this** they hope to raise wages of state employees.'
- (7) *Многие современные лекарственные препараты включают несколько ингредиентов, **благодаря чему** достигается большая эффективность.*
 'Many modern pharmaceuticals include several ingredients, **that's why** (lit. thanks to which) the greater effect is achieved'

Этом is used for inter-sentential connection, *эмо* - for intra-clausal connection. While the expressions *благодаря этому* 'due to this' and *благодаря чему* 'lit. due to what' are not fully grammaticalized, the expressions with the same core word

containing *то* ‘that’ (e.g. *благодаря тому, что* ‘due to the fact that’) is a complex conjunction included in dictionaries and grammars. There is also a parallel nominalized construction (as in *благодаря [его возвращению]_{EDU}* ‘due to his return’). This construction is a part of a finite clause. It constitutes an EDU. In this case, the connective can consist only of a core word and can be classified as a primary connective.

To sum up, the core word of some connectives can be combined with all of the mentioned above pronouns as well as with nominalized constructions. Thus, there are four possible constructions with the same meaning.

4.3. Comparison of Primary and Secondary Connectives

We conducted a comparative analysis of two classes of connectives, namely secondary vs. primary ones. The comparison is limited to the ‘Cause-Effect’ subset (141 example with overt connectives from 155 in total). The aim is to single out the features of these two classes that can help in further connectives extraction and classification.

Firstly, there is a statistically significant difference in the position of two types of connectives with respect to the satellite vs. nucleus opposition (a cause is a satellite and an effect is a nucleus in ‘Cause-Effect’ relation). Primary connectives are more often located in satellites, while secondary “prefer” nuclei (cf. Table 2, excluding 6 examples where connectives are in both EDUs). Therefore, primary connectives are located in an effect-span more frequently than secondary ones.

Table 2. The position of a connective in nucleus vs. satellite EDU with respect to connective type

connective type / position	nucleus	satellite	sum
primary	22	59	81
secondary	37	17	54
sum	59	76	135
$\chi^2(1) = 20.88, p < .001$ (Yates’ correction)			

The order of EDU differs, depending on the class of the connective used to signal the relation between EDUs. The preferable order is “nucleus-satellite” for relations marked with primary connectives and the satellite precedence is preferable with secondary ones (table 3, excluding 1 discontinuous EDU):

Table 3. The position of nucleus EDU in relation: primary vs. secondary connectives

connective type / span order	nucleus precedes	satellite precedes	sum
primary	44	38	82
secondary	19	39	58
sum	63	77	140
$\chi^2(1) = 5.18, p = .023$ (Yates’ correction)			

The primary connectives signal inter-clausal relations within a sentence more frequently than secondary ones, while the latter are used to mark inter-sentential relations or they are used for intra-clausal relations when discourse units are expressed via nominalizations:

Table 4. The difference between two classes of connectives wrt of signalling intra- vs. inter-clausal relations

connective type / type of connection	intra-clausal	inter-clausal	intra-sentential	sum
primary	61	11	11	83
secondary	26	19	13	58
<i>sum</i>	87	30	24	141
$\chi^2(2) = 12.34, p = .002$				

There is no statistically significant difference between spans size with secondary and primary markers.

5. Conclusion

In this paper, we present the analysis of discourse connectives in Russian. The analysis is based on detailed examination of ‘Cause-Effect’ connectives. In our research, we consider the connectives used for signaling the relations between text spans of different size (clause, sentence or bigger). We pay special attention to less grammaticalized constructions for Russian.

As a result, we suggest the list of causal relation connectives based on Ru-RSTreebank (it includes 48 elements), schemes for connectives annotation in the corpus and in the lexicon. All the examples of ‘Cause-Effect’ relation are annotated according to the corresponding scheme. The basic patterns for non-grammaticalised connectives, including verb and light verb constructions are determined. These patterns can be exploited to expand the list of causal connectives automatically via mining them in an unannotated corpus.

Besides, we provide the comparison of two classes of connectives. Our data show the statistically significant difference between primary vs. secondary connectives with respect to the nucleus/satellite position, intra- vs. inter-sentential relations and some others. The features that exhibit such a difference could be taken into consideration while developing machine learning technique for rhetorical relations extraction.

References

1. *Al-Saif A., Markert K.* (2010), The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic. In LREC 2010 Proceedings, pp. 2046–2053.
2. *Alonso L., Castellón I., Gibert K., Padró L.* (2002), An empirical approach to discourse markers by clustering. In Topics in Artificial Intelligence, Springer, Berlin, Heidelberg, pp. 173–183.
3. *Apresyan YU. D.* (2001), System-forming meanings 'to know' and 'to consider' in Russian [Sistemoobrazuyushchiye smysly «znat'» i «schitat'» v russkom yazyke], Russian Language and Linguistic Theory [Russkiy yazyk v nauchnom osheshchenii], 1, pp. 5–26.
4. *Asghar N.* (2016), Automatic Extraction of Causal Relations from Natural Language Texts: A Comprehensive Survey, arXiv preprint arXiv:1605.07895.
5. *Boguslavskaya, O. YU., Levontina, I. B.* (2004). Meanings 'cause' and 'purpose' in natural language [Smysly 'prichina' i 'tsel' v yestestvennom yazyke], Topics in the study of language [Voprosy yazykoznaniiya], (2), pp. 68–88.
6. *Carlson L., Marcu D.* (2001), Discourse tagging reference manual. ISI Technical Report ISI-TR-545, 54, 56.
7. *Chafe, W.* (1994), Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing, University of Chicago Press.
8. *Chang D. S., Choi K. S.* (2006), Incremental cue phrase learning and bootstrapping method for causality extraction using cue phrase and word pair probabilities, Information processing & management, Vol. 42, № 3, pp. 662–678.
9. *Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., ... & Schlaefel, N.* (2010), Building Watson: An overview of the DeepQA project. AI magazine, 31(3), 59–79.
10. *Galitsky B., Ilvovsky D. and Kuznetsov S. O.* (2018), Detecting Logical Argumentation in Text via Communicative Discourse Tree. To appear in Journal of Experimental and Theoretical Artificial Intelligence. JETAI 2018.
11. *Glovinskaya, M. Y.* (1993), Russian Speech Acts with the Meaning of Mental Pressure. / Logical analysis of language. Mental actions. Moscow: Science. [Russkie rechevihe akty so znacheniem mentaljnogo vozdeystviya / Logicheskij analiz yazihka. Mentaljnihe deystviya.]
12. *Grote B., Stede M.* (1998), Discourse marker choice in sentence planning, Proceedings of the 9th International Workshop on Natural Language Generation, Niagara-on-the-Lake, Canada, August 1998.
13. *Hirschberg, J., Litman, D.* (1993), Empirical studies on the disambiguation of cue phrases, Computational linguistics, 19(3), pp. 501–530.
14. *Khoo C. S. G. et al.* (1998), Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing, /Literary and Linguistic Computing, Vol. 13, № 4. pp. 177–186.
15. *Kibrik, A. A., Podlesskaya, V. I.* (2003), Towards building corpora of oral Russian speech: principles of transcription [K sozdaniyu korpusov ustnoy russkoy rechi: printsipy transkribirovaniya], Scientific and Technical Information Processing (series 2) Nauchno-tehnicheskaya informatsiya (seriya 2), 6, pp. 5–11.

16. Knott A. (1996), A data-driven methodology for motivating a set of coherence relations, Thesis (doctoral), University of Edinburgh.
17. Kustova, G. I. (2017), The Types of Adverbials [Tipy konstruksiy s adverbial-ami], Computational Linguistics and Intellectual Technologies: Proc. of the Int. Conf.» Dialogue, Vol. 2, pp.234–249.
18. Lee, A., Prasad, R., Webber, B., & Joshi, A. K. (2016), Annotating Discourse Relations with The PDTB Annotator. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations, pp. 121–125).
19. Louis, A., Joshi, A., & Nenkova, A. (2010), Discourse indicators for content selection in summarization. In Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue (pp. 147–156), Association for Computational Linguistics.
20. Mann W. C., Thompson S. A. (1988), Rhetorical Structure Theory: Toward a Functional Theory of Text Organization, Text 8, 3, pp. 243–281.
21. Meyer T., Webber B. *Implicitation of discourse connectives in (machine) translation* (2013), Proceedings of the Workshop on Discourse in Machine Translation, pp. 19–26.
22. Mírovský, J., Synková, P., Rysová, M., & Poláková, L. (2017), CzeDLex–A Lexicon of Czech Discourse Connectives. The Prague Bulletin of Mathematical Linguistics, 109(1), pp. 6191.
23. Paducheva, E. V. (2004), Dynamic models in the semantics of vocabulary [Dynamicheskiye modeli v yazyke]. M.: Languages of Slavonic Culture.
24. Pekelis, O. Ye. (2014), Causal subordinate clauses [Prichinnyye pridatochnyye], 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, 2015.
25. Pisarevskaya D. et al. (2017), Towards building a discourse-annotated corpus of Russian, Computational Linguistics and Intellectual Technologies: Proc. of the Int. Conf.» Dialogue, Vol. 1, pp. 194204.
26. Prasad R. et al. (2007), The penn discourse treebank 2.0 annotation manual.
27. Prasad R., Joshi A., Webber B. (2010), Realization of discourse relations by other means: alternative lexicalizations, Proceedings of the 23rd International Conference on Computational Linguistics: Posters, Association for Computational Linguistics, pp. 1023–1031.
28. Roze C., Danlos L., Muller P. (2012), LEXCONN: a French lexicon of discourse connectives, Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics, N^o. 10.
29. Rysová M., Rysova K. (2014), The Centre and Periphery of Discourse Connectives, PACLIC, pp. 452–459.
30. Schauer H. (2000), From elementary discourse units to complex ones, Proceedings of the 1st SIGdial workshop on Discourse and dialogue-Volume 10, Association for Computational Linguistics, pp. 46–55.
31. Shvedova, N. YU. ed. (1980), Russian grammar [Russkaya grammatika]. V dvukh tomakh. AN SSSR Institut russkogo jazyka. — M.: Nauka, 1980.

32. *Stede M., Umbach C. (1998), DiMLex: A lexicon of discourse markers for text generation and understanding, Proceedings of the 17th international conference on Computational linguistics-Volume 2, Association for Computational Linguistics, pp. 1238–1242.*
33. *Taboada, M., Voll, K., & Brooke, J. (2008), Extracting sentiment as a function of discourse structure and topicality. Simon Fraser University School of Computing Science Technical Report.*
34. *Toldova S. et al. (2017), Rhetorical relation markers in Russian RST Treebank, Proceedings of the 6th Workshop Recent Advances in RST and Related Formalisms, Santiago de Compostela, Spain, September 4 2017, Association for Computational Linguistics, pp. 29–33.*