

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2020»

Москва, 17–20 июня 2020 г.

ИМПЛИЦИТНЫЕ ЛОГИКО- СЕМАНТИЧЕСКИЕ ОТНОШЕНИЯ И МЕТОД ИХ ПОИСКА В ПАРАЛЛЕЛЬНЫХ ТЕКСТАХ

Гончаров А. А. (a.gonch48@gmail.com)

ИПИ ФИЦ ИУ РАН, Москва, Россия

Инькова О. Ю. (Olga.Inkova@unige.ch)

ИПИ ФИЦ ИУ РАН, Москва, Россия;

Женевский университет, Женева, Швейцария

DOI: 10.28995/2075-7182-2020-19-310-320

IMPLICIT LOGICAL-SEMANTIC RELATIONS AND A METHOD OF THEIR IDENTIFICATION IN PARALLEL TEXTS

Goncharov A. A. (a.gonch48@gmail.com)

Institute of Informatics Problems, FRC CSC RAS, Moscow,
Russia

Inkova O. Yu. (Olga.Inkova@unige.ch)

Institute of Informatics Problems, FRC CSC RAS, Moscow,
Russia; University of Geneva, Geneva, Switzerland

One of the main characteristics of logical-semantic relations (LSRs) between two fragments of a text is that these relations can be either explicit (expressed by some marker, e. g. a connective) or implicit (derived from the interrelation of these fragments' semantics). Since implicit LSRs do not have any marker, it is difficult to find them in a text (whether automatically or not).

In this paper, approaches to analysing implicit LSRs are compared, an original definition for them is offered and differences between implicit LSRs and LSRs expressed by non-prototypical means are described. A method is proposed to identify implicit LSRs using a parallel corpus and a supracorpora database of connectives. Based on the well-known statement that LSRs can be explicitated by adding connectives in the translation, it is argued here that through selecting pairs in which fragments where a connective is used to express an LSR in the translation correspond to those containing any of the translation stimuli standard for this connective in the source language, it is possible to get an array of contexts in which this LSR is implicit in the source text (or expressed by means other than connectives). This method is then applied to study the French causal connectives *car*, *parce que* and *puisque* using a Russian-French parallel corpus. The corpus data are analysed to obtain information about LSRs particularly about cases where the causal LSR in Russian is implicit, as well as about the use of causal connectives in French. These results are used to show that the method proposed allows to quickly create a representative array of contexts with implicit LSRs, which can be useful in both text analysis and in machine learning.

Key words: corpus linguistics, parallel corpora, logical-semantic relations, implicitness, discourse connectives

1. Введение

Под логико-семантическими отношениями (ЛСО) мы понимаем такие отношения между двумя фрагментами текста, которые потенциально могут быть выражены специализированным показателем, коннектором (ср. отношения причины, следствия, сопоставления и др., выражаемые союзами). Уже тот факт, что в этом определении фигурирует слово «потенциально», свидетельствует о том, что ЛСО могут быть как эксплицитными, т. е. выраженными соответствующим показателем, так и имплицитными, т. е. выводимыми на основе смыслового соотношения следующих друг за другом фрагментов текста, коммуникативной ситуации или знаний об устройстве мира. Если коннекторы и, соответственно, эксплицитные ЛСО довольно хорошо описаны, то в изучении имплицитных ЛСО пока остаётся много белых пятен. Специального рассмотрения заслуживает вопрос об их идентификации в тексте, что становится особенно актуальным при разработке систем автоматического анализа и генерации текста. Как известно, уровень консенсуса при разметке имплицитных отношений гораздо ниже, чем при разметке эксплицитных, а сами разметчики часто опираются в своих решениях на возможность подстановки того или иного коннектора между двумя высказываниями. Кроме того, данных о частотности случаев, когда то или иное ЛСО является имплицитным, практически не существует, что обусловлено сложностью получения представительного массива аннотаций с имплицитными отношениями.

Мы начнём с обзора существующих подходов к исследованию имплицитных ЛСО и их определения, после чего опишем разработанный нами метод их поиска с использованием параллельных текстов и надкорпусной базы данных (НБД) коннекторов, а также результаты его применения.

2. Имплицитные ЛСО: подходы и определения

Имплицитные отношения между двумя следующими друг за другом фрагментами текста признаются всеми без исключения существующими теориями анализа текста и дискурса, а также грамматиками. Однако содержание понятия имплицитного отношения в значительной степени различается от подхода к подходу. В грамматиках имплицитные ЛСО (т. е. в данном случае те, которые потенциально могут быть выражены сочинительными и подчинительными союзами или относительными словами) рассматриваются в разделе, посвящённом бессоюзной синтаксической связи (ср. для русского языка [Виноградов, Истрина 1960а]; [Виноградов, Истрина 1960б]; [Шведова 1970]; [Шведова 1980], аналогичная ситуация наблюдается и в грамматиках основных европейских языков). В то же время отмечается, что «круг отношений, выражаемых бессоюзными сложными предложениями, не совпадает с соответствующими функциями сложносочинённых и сложноподчинённых предложений» [Виноградов, Истрина 1960а: 102]. Основным критерием для выделения данного типа сложных предложений служит, тем не менее, отсутствие показателя синтаксической связи между его частями, о чём говорит его название.

В наиболее известных теориях анализа структуры текста — Теории риторической структуры (RST) У. Манна и С. Томпсон, Теории сегментной репрезентации дискурса (SDRT) Н. Ашера и теоретических положениях, лежащих в основе разработки Пенсильванского аннотированного корпуса (PDTB), — признаётся возможность имплицитных отношений между дискурсивными единицами (ДЕ). Но сами термины «отношение» и, следовательно, «имплицитное отношение» понимаются по-разному.

RST и SDRT можно определить как теории, основанные на «глобальном» подходе к связности текста. Это означает, что текст представляет собой единый дискурсивный объект, все элементы которого связаны между собой «риторическими» (в терминах RST) отношениями. Следовательно, понятие «риторическое» отношение является родовым по отношению к понятию ЛСО (подробнее см. [Инькова 2019а]), а коннектор рассматривается лишь как один из видов показателей риторических отношений, причём как показатель малозначимый, поскольку его функция заключается только в том, чтобы сузить спектр возможных интерпретаций отношений между ДЕ [Mann, Thompson 1986: 71]. При этом допускается, что риторическое отношение может не иметь никакого показателя [Там же: 68]. Однако, поскольку в литературе, посвящённой маркированию отношений, коннекторы долгое время считались единственными показателями отношений связности, немаркированными (имплицитными) считались только такие отношения, которые не выражены коннектором; см. [Martin 1992]; [Renkema 2004]; [Taboada 2009]. Тем не менее, как показано в [Taboada 2006], коннекторы маркируют только 30–40% (в зависимости от типа текста, письменного vs. устного) риторических отношений. Поэтому вопрос о средствах маркирования отношений был пересмотрен, в частности в связи с применением RST для создания аннотированных корпусов текстов (см. [Carlson et al. 2003]) и автоматической идентификации риторических отношений. К коннекторам добавляются показатели самой разнообразной природы: лексические,

морфологические (временные формы), семантические (синонимия, антонимия и др.), синтаксические (различные виды придаточных и др.), графические (знаки препинания и др.) и т. д.; они могут также комбинироваться между собой [Das, Taboada 2013: 5–6]; [Taboada, Das 2013: 257]; [Das, Taboada 2014: 8–9]. Сам термин «имплицитное» отношение исчезает из пособий по аннотированию; см., например, [Carlson, Marcu 2001]; [Das, Taboada 2014]. В случае, когда никакой из потенциальных показателей отношения не может быть идентифицирован, проставляется метка *unsure*; ср. (1), приводимый в [Das, Taboada 2014: 33]:

- (1) {*This hasn't been Kellogg Co.'s year.*} {*The oat-bran craze has cost the world's largest cereal maker market share.*} Cause [wsj_610: 1/2]¹

В SDRT, также характеризующейся глобальным подходом к связности текста [Asher 1993], признаётся, что показателями дискурсивных отношений являются коннекторы. Таблица «показателей отношений» в [Muller et al. 2012: 27–28] содержит только единицы данного класса, причём подчёркивается, что коннекторы — многозначны и что указанные для них отношения (одно или более для каждого показателя) являются скорее ориентиром для разметчиков.

PDTV отличает «локальный» подход к связности текста (подробнее о различиях в аннотировании между RST и PDTV см. [Poláková et al. 2017]). Этот корпус задумывался с целью аннотировать коннекторы и выражаемые ими отношения [Prasad et al. 2006: 1]. Коннекторы подразделяются на эксплицитные и имплицитные. В последнем случае разметчик должен попытаться понять, какое отношение может быть установлено между двумя фрагментами текста, а аннотирование состоит в том, чтобы вставить в текст наиболее подходящий по смыслу коннектор. Так, в (2), согласно [Prasad et al. 2007: 17], между фрагментами текста могут быть вставлены два коннектора: *because* и *for example*, потому что второй фрагмент текста вводит одно из объяснений утверждения «The third principal in the S. Gardens adventure did have garden experience»:

- (2) {*The third principal in the S. Gardens adventure did have garden experience.*}
{(Импл. = **because for example**) *The firm of Bruce Kelly/David Varnell Landscape Architects had created Central Park's Strawberry Fields and Shakespeare Garden.*}

Кроме того, PDTV проводит различие между имплицитным коннектором (*implicit connective*) и имплицитным отношением (*implicit relation*), поскольку случаи, когда между двумя фрагментами текста невозможно вставить коннектор разделяются на 1) альтернативные лексикализации (AltLex), когда отношение выражено другими лексическими средствами, традиционно не относящимися к классу коннекторов, а подстановка коннектора приводит к избыточности высказывания; 2) случаи, когда между фрагментами существует недискурсивное отношение, а именно, когда второй фрагмент текста описывает некоторый аспект ситуации, описанной в первом фрагменте текста (EntRel,

¹ Во всех примерах вне зависимости от исходного оформления связываемые фрагменты заключены в фигурные скобки, а языковые единицы, выражающие ЛСО, выделены полужирным шрифтом.

entity-based coherence [Knott et al. 2001]); 3) случаи, когда между фрагментами текста не может быть установлено никакого отношения (NoRel).

Аннотации имплицитных коннекторов и контекстов, относящихся к одному из трёх перечисленных выше случаев, изначально объединялись в группу «implicit relation annotations» [Prasad et al. 2006: 1], однако затем эксплицитным дискурсивным отношениям были поставлены в соответствие 1) эксплицитные коннекторы и 2) альтернативные лексикализации, а имплицитным дискурсивным отношениям — только случаи с имплицитным коннектором [Prasad et al. 2017: 1200–1201].

В разрабатываемом нами подходе и в НБД коннекторов² под имплицитными ЛСО понимаются такие отношения между двумя или более фрагментами текста, которые не выражены ни коннектором, ни каким-либо другим языковым средством (лексическим или грамматическим), ни семантически нагруженным знаком препинания (прежде всего, тире и двоеточием; подробнее см. [Инькова 2019б]), но которые могут быть выражены коннектором без необходимости модифицировать контекст. Последнее положение отличает наш подход от подхода PDTB.

Так, для нас в (3) отношение причины во фрагменте на русском языке выражено деепричастным оборотом, для которого традиционно выделяется это значение [Виноградов, Истрина 1960а: 650–656]; [Шведова 1970: 645]; [Шведова 1980: 181–182].

- (3) {Вдруг Штольц изменился в лице,} {**поймав** истину.}
[Гончаров И. А. Обломов (1848–1859)]

В то же время в PDTB аналогичный случай (4) [Webber et al. 2019: 34] относится к случаям с «имплицитным коннектором», и при аннотировании вводится коннектор причины *as result of*, что приводит к превращению формы present participle в форму gerund и к семантической избыточности, поскольку для подобных конструкций в английском языке также выделяется значение причины (см., например, [Stump 1985], где, в частности, приводится таблица, показывающая, что причинная семантика данных единиц выделяется во всех рассматриваемых автором работах по грамматике английского языка [Там же: 19]).

- (4) {*Treasurys opened lower,*} {(Импл. = **as a result of**) *reacting negatively to news that the producer price index — a measure of inflation on the wholesale level — accelerated in September.*} [wsj 2428]

При этом критерием, на основании которого в PDTB выделена группа AltLex, является как раз возникновение избыточности в случае добавления коннектора ([Prasad et al. 2007: 27]; [Webber et al. 2019: 12])³.

² Представительный фрагмент НБД коннекторов доступен по адресу: <http://a179.frccsc.ru/RSCF10004/main.aspx>.

³ Заметим, что данный подход ориентирован на английский язык; ср. невозможность такой трансформации в русском и многих других языках: (4₁) {Рынок казначейских бумаг открылся снижением,} {(Импл. = *как результат / *в результате / *вследствие) негативно отреагировав на новости о том, что индекс цен производителей — показатель инфляции на оптовом уровне — в сентябре вырос.} При возможном: **вследствие** негативной реакции...

Наше понимание имплицитного ЛСО является, таким образом, более узким, чем принятое в PDTB, наиболее близком нам подходе, соответствуя только части случаев с «имплицитным коннектором».

3. Метод поиска имплицитных ЛСО в параллельных текстах

Плодотворность использования параллельных текстов для аннотирования ЛСО, как эксплицитных, так и имплицитных, уже не раз отмечалась исследователями. Так, в [Cartoni et al. 2013] показано, что переводной эквивалент коннектора может дать дополнительную информацию о его семантике и снять неоднозначность интерпретации в случае полисемичных коннекторов. В [Shi et al. 2017: 484] подчёркивается, насколько трудным является распознавание имплицитных ЛСО и что нейронные сети плохо справляются с этой задачей из-за нехватки аннотированных данных. Для решения этой практической проблемы в работе предлагается оригинальный метод использования параллельных корпусов (в данном случае англо-французского корпуса). Авторы исходят из положений о том, что профессиональные переводчики иногда добавляют коннекторы, отсутствующие в оригинале (т.н. эксплицитирование ЛСО), и что при машинном переводе уже имеющиеся в тексте коннекторы, как правило, не опускаются. Поэтому в рамках исследования выполняется обратный перевод текстов с французского на английский язык при помощи системы машинного перевода. После этого отбираются случаи, когда ЛСО в оригинальном фрагменте на английском языке является имплицитным, а во фрагменте, полученном при помощи обратного машинного перевода на английский язык, оно выражено коннектором — из-за того, что соответствующий коннектор был добавлен профессиональными переводчиками при переводе оригинального фрагмента на французский язык. Это позволяет, по мнению авторов, получить информацию высокой степени достоверности о ЛСО, содержащемся в исходном английском тексте, и создать представительный массив аннотаций для машинного обучения.

В работе [Гончаров, Инькова 2019а] был предложен иной метод поиска имплицитных ЛСО в параллельных текстах. В общем случае, если корпус таких текстов позволяет осуществлять поиск языковых единиц (ЯЕ) как в оригинале, так и в переводе, то можно находить 1) пары фрагментов, где в одном из языков присутствует некоторая ЯЕ; 2) пары фрагментов, где в одном языке употреблена ЯЕ x , а в другом языке — ЯЕ y . В первом случае нередко приходится работать с большим массивом случаев, во втором имеется возможность задавать лишь известные (большой частью стандартные) комбинации x и y .

Исходя из этого, представляется полезным сократить массив таким образом, чтобы в нём остались лишь неизвестные или нестандартные соответствия. Для этого в НБД была предусмотрена возможность поиска, позволяющая отбирать пары фрагментов, где во фрагменте на одном языке обязательно должна присутствовать заданная ЯЕ, а во фрагменте на другом языке должна/ы отсутствовать заданная/ые ЯЕ. На первом этапе работы была реализована лишь функция исключения одной или нескольких однословных ЯЕ [Гончаров, Инькова 2019а]. Эта функция показала эффективность такого вида поиска,

поэтому ввиду формальных особенностей коннекторов, которые часто являются неоднословными (51,25% аннотаций в НБД на 15.02.2020), она была усовершенствована, что дало возможность исключать из выборки одну или несколько неоднословных единиц в фиксированной форме.

4. Анализ полученных результатов

Для настоящего исследования были использованы тексты параллельного французского подкорпуса НКРЯ (направление перевода «русский — французский») общим объёмом более 3,9 млн словоупотреблений (данные на 15.02.2020). Поиск был применён для анализа русских стимулов перевода трёх французских коннекторов — *car, parce que, puisque*, — выражающих ЛСО причины. Из выборки исключались пары фрагментов, которые содержат коннекторы русского языка *потому (что), оттого (что), поскольку, так как, ибо, за то что, ведь*. Безусловно, не во всех случаях стимул перевода был нулевым (ср. (3) выше), однако в данной работе мы остановимся лишь на таких примерах, которые мы объединяем в кластер Zero. Этот кластер в ходе анализа выборки был разделён на три подкластера:

1. Коннектор отсутствует, но может быть восстановлен (Absent CNT).

- (5) {Совещанию придавалось особое значение,} {(Импл. = **поскольку**)
она связывалось с поездкой Гиммлера в полевую Ставку вождя.}
[Гроссман В. С. Жизнь и судьба (1960)]
{Cette réunion était particulièrement importante,} {**puisque** Himmler devait
se rendre ensuite au Quartier général du Führer.} [Пер. Berelowitch A. (1980)]

2. Фрагмент, соответствующий фрагменту с коннектором, имеет иную структуру (DifferStr FragmCNT); так, в (6) при переводе фрагмент текста, содержащий коннектор, меняет свою структуру: ‘никто не хочет брать меня на работу, потому что я из Чечни’.

- (6) {На работу не хотят брать,} {“А-а-а... из Чечни?”}
[Алексиевич С. А. Время секунд хэнд (ч. 2) (2013)]
{Personne ne veut m'employer} {**parce que** je reviens de Tchétchénie...}
[Пер. Benesch S. (2013)]

3. Соответствующий фрагмент с коннектором отсутствует (Absent FragmCNT); ср. (7), где переводчик вводит дополнительный фрагмент ‘вам же хуже’ (*tant pis pour vous*), с которым он связывает отношением причины фрагмент ‘вам же нечего будет в пирожное положить’. В оригинале фрагменты текста *просыплю корицу и вам же нечего будет в пирожное положить* связаны ЛСО условия.

- (7) — Смотрите, {просыплю корицу,} {вам же нечего будет в пирожное положить,} — заметила она. [Гончаров И. А. Обломов (1848–1859)]
— Attention, je vais renverser la cannelle; {tant pis pour vous} {**car** je n'aurai rien à mettre dans votre gâteau,} fit-elle remarquer. [Пер. Jurgenson L. (1988)]

Несмотря на то, что изменения контекста при переводе, приводящие к соответствиям типа 2. и 3., могут быть оправданными, мы не рассматриваем их как случаи имплицитных ЛСО. Тем не менее, в [Shi et al. 2017: 491] особо отмечается важность исследования контекстов, сходных с теми, которые мы относим к типу 2. Хотя такие контексты 1) не включают никакого коннектора, 2) имеют структуру, не позволяющую его добавить, и 3) могут не рассматриваться как содержащие какое-либо ЛСО, они вызывают появление коннектора в переводе, а потому могут дать новую информацию о механизмах функционирования ЛСО в целом и расширить имеющиеся знания о спектре контекстов, включающих ЛСО.

Таблица ниже содержит данные о количестве переводных соответствий, относящихся к кластеру Zero, и их распределение по трём подкластерам (использование различных речевых реализаций коннекторов в таблице не отражается, т. е. случаи, например, с *car*, *car enfin*, *car...* et объединены в кластер *car*). Все процентные показатели приведены по отношению к общему числу пар с этим коннектором в НБД.

Коннектор франц. яз.	Стимул перевода в русском языке								Всего пар в НБД
	Zero		Absent CNT		DifferStr FragmCNT		Absent FragmCNT		
	Абс.	%	Абс.	%	Абс.	%	Абс.	%	
<i>car</i>	178	18,46	148	15,35	21	2,18	9	0,93	964
<i>parce que</i>	44	7,84	34	6,06	10	1,78	0	0	561
<i>puisque</i>	18	9,78	12	6,52	5	2,72	1	0,54	184
	240		194		36		10		1709

Таблица показывает, прежде всего, эффективность предлагаемого метода поиска: подавляющее большинство полученных аннотаций содержат именно контексты с имплицитным ЛСО причины в русском оригинале, т. е. когда коннектор отсутствует, но может быть восстановлен. Во-вторых, можно заметить, что изменение структуры фрагмента текста с коннектором или добавление фрагмента текста в перевод не являются частотными. Наконец, из таблицы видно, что причинные коннекторы имеют различную дистрибуцию: для коннектора *car* нулевой стимул перевода встречается гораздо чаще, чем для других коннекторов причины. На наш взгляд, это вызвано тем, что данный коннектор выражает более «слабое» причинное отношение, которое в оригинале установлено, как правило, между самостоятельными предложениями вне зависимости от их объёма; клаузами в составе сложного предложения, разделёнными запятой или точкой с запятой; частицами «да» или «нет», выражающими согласие или несогласие, и причинами этих ответов; частями одной и той же реплики, разделёнными авторской речью.

5. Выводы

Описанный вид поиска позволяет оперативно получить представительные массивы случаев, когда ЛСО устанавливается без использования прототипических средств. Аннотирование этих случаев даёт возможность получить количественные

показатели средствами НБД, а дальнейший анализ позволяет 1) классифицировать случаи неспециализированного выражения ЛСО, играющих, как известно, ведущую роль в обеспечении связности текста (такая классификация является одной из задач теоретической и когнитивной семантики); 2) классифицировать случаи, когда ЛСО является имплицитным, что может быть использовано при разработке систем автоматической обработки текста, в том числе его семантического анализа; 3) сравнить эффективность существующих методов поиска имплицитных ЛСО. В данной работе было предложено решение только для второй задачи. О методах решения первой задачи см., например, [Инькова 2019б]; [Гончаров, Инькова 2019б]), где предложена в общем виде и проиллюстрирована типология средств выражения ЛСО, альтернативных коннекторам.

Достоинством предлагаемого подхода является как возможность рассмотрения конкретных языковых единиц (насколько часто стимул перевода конкретного коннектора является нулевым), так и — впоследствии — ЛСО в целом (насколько часто случаи имплицитного ЛСО эксплицируются в переводе; какие ЛСО чаще других являются имплицитными). Привлечение данных русского и французского языков позволит верифицировать гипотезы, касающиеся общих принципов функционирования языка, сформулированные на основе данных английского языка, в частности «causality-by-default hypothesis» [Sanders 2005]. Отметим также, что в работах, содержащих описание методов автоматизации поиска имплицитных ЛСО, степень детализации результатов гораздо ниже, поскольку случаи имплицитных ЛСО распределяются по группам отношений, а их классификации, как известно, в значительной степени не совпадают.

Перспективами исследования является получение количественных данных по коннекторам, выражающим другие ЛСО, что позволит в дальнейшем перейти от исследования случаев, когда ЛСО выражено коннектором в оригинале или в переводе, к целенаправленному поиску случаев имплицитных ЛСО, не ограниченному использованием параллельных корпусов.

Литература

1. Asher N. (1993) Reference to abstract objects in discourse. Kluwer Academic Publishers, Dordrecht.
2. Carlson L., Marcu D. (2001) Discourse Tagging Reference Manual. ISI Technical Report ISI-TR-545, available at: <https://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>.
3. Carlson L., Marcu D., Okurowski M. E. (2003) Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory, Current and New Directions in Discourse and Dialogue, Kluwer Academic Publishers, Dordrecht, pp. 85–112.
4. Cartoni B., Zufferey S., Meyer T. (2013) Annotating the Meaning of Discourse Connectives by Looking at their Translation: The Translation Spotting Technique, Dialogue and Discourse, Vol. 4, No. 2, pp. 65–86.
5. Das D., Taboada M. (2013). Explicit and Implicit Coherence Relations: A Corpus Study, Proceedings of the 2013 annual conference of the Canadian Linguistic

- Association, available at: http://homes.chass.utoronto.ca/~cla-acl/actes2013/Das_and_Taboada-2013.pdf.
6. *Das D., Taboada M.* (2014) RST Signalling Corpus Annotation Manual, available at: https://www.sfu.ca/~mtaboada/docs/publications/RST_Signalling_Corpus_Annotation_Manual.pdf.
 7. *Goncharov A. A., Inkova O. Yu.* (2019a) Methods for identification of implicit logical-semantic relations in texts [Metodika poiska implitsitnykh logiko-semanticheskikh otnosheniy v tekste], Informatics and Applications [Informatika i ee Primeneniya], Vol. 13, No. 3, pp. 97–104.
 8. *Goncharov A. A., Inkova O. Yu.* (2019b) Means of Expressing Causal Relations in Russian: Analysis Using a Cross-Linguistic Supracorpora Database [Sposoby vyrazheniya prichinnykh otnosheniy v russkom yazyke: opyt analiza s ispol'zovaniem krosslingvisticheskoy nadkorpusnoy bazy dannykh], Russian Grammar: Active Processes in Language and Discourse. Proceedings of International Scientific Symposium [Russkaya grammatika: aktivnye protsessy v yazyke i rechi. Sbornik nauchnykh trudov Mezhdunarodnogo nauchnogo simpoziuma], Yaroslavl, pp. 385–396.
 9. *Inkova O. Yu.* (2019a) Logical-semantic relations: classification problems [Logiko-semanticheskie otnosheniya: problemy klassifikatsii], Text coherence: mereological logical-semantic relations [Svyaznost' teksta: mereologicheskie logiko-semanticheskie otnosheniya], Izdatel'skii Dom YaSK, Moscow, pp. 11–98.
 10. *Inkova O. Yu.* (2019b) Annotation of Parallel Texts: The Concept of Divergent Translation [Annotirovanie paralel'nykh tekstov: ponyatie «divergentnyy perevod»], Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue” [Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoy mezhdunarodnoy konferentsii «Dialog»], Moscow, pp. 227–238.
 11. *Knott A., Oberlander J., O'Donnell M., Mellish C.* (2001) Beyond elaboration: the interaction of relations and focus in coherent text, Text Representation: Linguistic and Psycholinguistic Aspects, John Benjamins, Amsterdam, pp. 181–196.
 12. *Mann W. C., Thompson S. A.* (1986) Relational Propositions in Discourse, Discourse Processes, Vol. 9, pp. 57–90.
 13. *Martin J. R.* (1992) English Text: System and Structure, John Benjamins, Amsterdam and Philadelphia.
 14. *Muller Ph., Vergez-Couret M., Prévot L., Asher N., Benamara F., Bras M., Le Draoulec A., Vieu L.* (2012) Manuel d'annotation en relations de discours du projet AN-NODIS, Carnets de Grammaire 21. Rapports internes de CLLE-ERSS, available at: <http://w3.erss.univ-tlse2.fr/textes/publications/CarnetsGrammaire/carnGram21.pdf>.
 15. *Poláková L., Mírovský J., Synková P.* (2017) Signalling Implicit Relations: A PDTB — RST Comparison, Dialogue & Discourse, Vol. 8, No. 2, pp. 225–248.
 16. *Prasad R., Miltsakaki E., Dinesh N., Lee A., Joshi A., Robaldo L., Webber B. L.* (2007) The Penn Discourse Treebank 2.0 Annotation Manual. IRCS Technical Reports Series, available at: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1203&context=ircs_reports.

17. Prasad R., Miltsakaki E., Dinesh N., Lee A., Joshi A., Webber B. L. (2006) The Penn Discourse TreeBank 1.0 Annotation Manual. IRCS Technical Reports Series, available at: https://repository.upenn.edu/cgi/viewcontent.cgi?referer=https://www.google.fr/&httpsredir=1&article=1003&context=ircs_reports.
18. Prasad R., Webber B., Joshi A. (2017) The Penn Discourse Treebank: An Annotated Corpus of Discourse Relations, Handbook of Linguistic Annotation, Springer Science+Business Media, Dordrecht, pp. 1197–1217.
19. Renkema J. (2004) Introduction to Discourse Studies, John Benjamins Publishing Company, Amsterdam / Philadelphia.
20. Sanders T. J. M. (2005) Coherence, causality and cognitive complexity in discourse, Proceeding SEM-05, First International Symposium on the Exploration and Modelling of Meaning, pp. 105–114.
21. Shi W., Yung F., Rubino R., Demberg V. (2017) Using Explicit Discourse Connectives in Translation for Implicit Discourse Relation Classification, Proceedings of The 8th International Joint Conference on Natural Language Processing, Taipei, pp. 484–495.
22. Stump G. T. (1985) The Semantic Variability of Absolute Constructions, D. Reidel Publishing Company, Dordrecht.
23. Taboada M. (2006) Discourse markers as signals (or not) of rhetorical relations, Journal of Pragmatics, Vol. 38, No. 4, pp. 567–592.
24. Taboada M. (2009) Implicit and explicit coherence relations, Discourse, of Course, John Benjamins Publishing Company, Amsterdam, pp. 127–140.
25. Taboada M., Das, D. (2013). Annotation upon annotation: Adding signalling information to a corpus of discourse relations, Dialogue and Discourse, Vol. 4, No. 2, pp. 249–281.
26. Webber B., Prasad R., Lee A., Joshi A. (2019) The Penn Discourse Treebank 3.0 Annotation Manual, available at: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.
27. Vinogradov V. V., Istrina E. S. (eds.) (1960a) Grammar of Russian Language. Vol. II. Syntax. Part 1 [Grammatika russkogo yazyka. T. II. Sintaksis. Ch. 1.], Izdatel'stvo AN SSSR, Moscow.
28. Vinogradov V. V., Istrina E. S. (eds.) (1960b) Grammar of Russian Language. Vol. II. Syntax. Part 2 [Grammatika russkogo yazyka. T. II. Sintaksis. Ch. 2.], Izdatel'stvo AN SSSR, Moscow.
29. Shvedova N. Yu. (ed.) (1970) Grammar of Modern Standard Russian Language [Grammatika sovremennogo russkogo literaturnogo yazyka], Nauka, Moscow.
30. Shvedova, N. Yu. (ed.) (1980) Russian Grammar. Vol. II. Syntax [Russkaya grammatika. Tom II. Sintaksis.], Nauka, Moscow.