

23–25 апреля 2025 г.

Multi-senses connectives: some precisions

Inkova O. Yu.

Institute of Informatics Problems, FRC CSC RAS,
Moscow, Russia
University of Geneva, Geneva, Switzerland
Olga.Inkova@unige.ch

Abstract

The article discusses multi-senses connectives and their annotation in text corpora. The author clarifies the concept, which is usually applied to three different phenomena: i) the uncertainty of annotators in choosing one of the meanings of a polysemic connector; ii) the possibility of establishing more than one relation, both explicit and implicit, between two fragments of text; iii) the “combination” of several meanings by a connective. The author then considers how these cases are annotated in the Penn Discourse Treebank and in the Supracorpora Database of Connectives, which use a multi-label annotation for discourse relations. The study also raises a number of theoretical questions: what information can we obtain from double labels, which relations are distributionally close, i.e. can be established in the same contexts, how to separate the contribution of the connective and of the context to the overall meaning of a sequence of sentences, how legitimate is it to talk about the combination of meanings by the connective, how to annotate the features of the context of the connectives so that this data can be used for Natural Language Processing? The solution to these questions is important for both theoretical cognitive and applied research (in particular, for machine learning and machine translation).

Keywords: semantics; discourse relations; supracorpora database; connectives; corpus linguistics

DOI: 10.28995/2075-7182-2025-23-129-136

Неоднозначность коннекторов: некоторые уточнения

Инькова О. Ю.

ИПИ ФИЦ ИУ РАН, Москва, Россия
Женевский университет, Женева, Швейцария
Olga.Inkova@unige.ch

Аннотация

В статье рассматриваются неоднозначные коннекторы и их разметка в корпусах текстов. Автор уточняет понятие, которое применяется, как правило, для трех разных явлений: i) колебания разметчиков в выборе одного из значений полисемичного коннектора; ii) возможность установить между одинаковыми фрагментами текста более одного отношения, как эксплицитного, так и имплицитного; iii) «совмещение» коннектором нескольких значений. Далее рассматривается, как эти случаи размечаются в Пенсильванском аннотированном корпусе и в Надкорпусной базе данных коннекторов, в которых используется аннотирование с несколькими метками для дискурсивных отношений. Исследование ставит также ряд теоретических вопросов: какую информацию мы можем извлечь из комбинированных меток, какие отношения дистрибутивно близки, т.е. могут быть установлены в одинаковых контекстах, как различить вклад коннектора и контекста в общее значение последовательности высказываний, насколько правомерно говорить о совмещении значений коннектором, как размечать особенности контекста употребления коннектора, чтобы эти данные могли быть использованы для автоматической обработки текста? Решение этих вопросов важно как для теоретических когнитивных, так и для прикладных исследований (в частности, для машинного обучения и машинного перевода).

Ключевые слова: семантика; дискурсивные отношения; надкорпусная база данных; коннекторы; корпусная лингвистика

1 Введение

Семантика коннекторов не перестает привлекать внимание лингвистов и специалистов по автоматической обработке текстов и машинному обучению. Об этом, в частности, свидетельствует свежая монография (Zufferey & Degand 2024), в которой, среди прочих вопросов, связанных с коннекторами и выражаемыми ими отношениями, обсуждается вопрос об их неоднозначности. Сам по себе этот вопрос не вызывает сомнения, но его постановка требует уточнения. Действительно этот термин используется для трех разных явлений. В первом случае речь идет о выборе одного из значений полисемичного коннектора, что отражается, в частности, в колебаниях разметчиков корпусов текстов в выборе той или иной метки (Stede 2008; Spooren & Degand 2010; Webber 2013). Иными словами, в основе данного вида неоднозначности лежит возможность множественной интерпретации одного элемента, в нашем случае коннектора. Например, англ. *so* может выражать как отношение Результат, так и отношение Цель; ср. (1) и (2), приводимые в (Prasad et al. 2018: 46¹):

- (1) *Most murders are state crimes so any federal capital-punishment law probably would turn out to be more symbolism than substance* [wsj 0426]
- (2) *Rep. Oakar, a Democrat from Cleveland, wants a \$6.9 million grant so Cleveland can build an 18-story Rock and Roll Hall of Fame* [wsj 1847]

В последнее время этот вопрос активно обсуждается и для имплицитных отношений, т. е. тех случаев, когда коннектор между фрагментами текста отсутствует, но может быть ‘восстановлен’ (Scholman et al. 2022; López Cortez & Jacobs 2023; Long et al. 2024). Так, в примере (3) четыре разметчика оценили отношение между вторым и третьим предложениями как Соединительное, а пять – как Результат.

- (3) *It is logical that our attention is focused on cities. *Cities are home to 80% of the 500 million or so inhabitants of the EU. It is in cities that the great majority of jobs, companies and centres of education are located.** (пример из López Cortez & Jacobs 2023: 1)

Во втором случае под неоднозначностью понимается возможность установить между двумя одинаковыми фрагментами текста более одного дискурсивного отношения. Речь идет о многокомпонентных коннекторах, каждая из составляющих которых сохраняет свою семантику. Например, в (4)² в составе многоэлементного коннектора *a... тем временем* союз *a* выражает сопоставительные отношения, а *тем временем* – одновременность:

- (4) В руках он держал книгу. За ним следом – молодой дьяк. Заметив краем глаза, что справедливость вот-вот будет восстановлена, Виктор продолжил свой путь. *А дьяк тем временем вырвал из рук мальчишки книгу, ударил этой книгой его несколько раз по голове и тоже поспешил обратно к церкви.* [Андрей Курков. Закон улитки (2005)]

Наконец, в некоторых исследованиях утверждается, что коннекторы могут выражать одновременно несколько отношений. Так, например, в работе (Zufferey, Degand 2024: 147) указано, что фр. *tant que* выражает одновременно временное и условное значения, а в работе (Кобозева и др. 2023: 41) – что русск. *в то же время* совмещает временное и противительное значения; в (Prasad et al. 2018: 52) приводится следующий пример на совмещение отношений временного следования и причины с соответствующей разметкой.

- (5) *I cannot recall any disorder in currency markets since (SUCCESSION,REASON) the 1974 guidelines were adopted.* [wsj 0779]

Важность адекватного решения вопроса неоднозначности коннекторов для аннотирования и автоматической обработки текстов, а также для улучшения качества машинного перевода³ трудно переоценить. С этим связаны также и более фундаментальные вопросы: является ли отношения,

¹ Мы сохраняем форму представления примеров.

² При отсутствии других указаний и используются примеры из Национального корпуса русского языка (НКРЯ).

³ О трудностях машинного перевода коннекторов, особенно состоящих из более одного элемента, см. (Nuriev et al. 2018), где содержится также обширная библиография по данному вопросу.

выражаемые коннектором семантически близкими; какие отношения могут сочетаться, а какие нет; какие отношения имеют одинаковую контекстную дистрибуцию; как распределяются роли коннектора и контекста в общей интерпретации последовательности высказываний?

2 Разметка неоднозначных коннекторов в аннотированных корпусах текстов

Хорошо известно, что аннотирование дискурсивных отношений представляет значительные трудности. Поэтому разметка во многих корпусах опирается на семантику показателя, и пособия по аннотированию содержат списки коннекторов с указанием выражаемых отношений. Так, например, разметчикам Пенсильванского аннотированного корпуса (PDTB) в случае эксплицитных отношений, т.е. маркированных коннектором, предлагается выбрать из списка соответствующее коннектору отношение, а в случае имплицитных отношений, как видно в (8) ниже, предлагается подставить показатель, подходящий по смыслу, и на этом основании проставить соответствующее отношение (ср. PDTB 2008; Prasad et al. 2019). Проблема состоит в том, что многие коннекторы, как мы видели, неоднозначны, и разметчики не всегда могут однозначно определить выражаемое ими отношение, а кроме того, отношений между аннотируемыми фрагментами текста может быть несколько. В качестве решения данной проблемы предлагается разметка типа multi-label Annotation, или аннотирование с несколькими метками. Однако, хотя в последнее время во многих работах по аннотированию, причем в разных областях (ср. Aroyo & Welty 2015; Basile et al. 2021; Marchal et al. 2022), признается, что разметчики могут не соглашаться в проставлении метки и что при этом каждый из них может быть прав в силу субъективности и сложности многих задач, в аннотированных корпусах текстов обычно отсутствуют аннотации более чем одного дискурсивного отношения между двумя одинаковыми фрагментами текста. Например, в корпусах, использующих Теорию риторической структуры (RST), допускается альтернативный анализ дискурсивной структуры текста, но предлагается выбирать одно отношение, а именно то, которое соответствует коммуникативным целям автора (Mann et al. 1989; Carlson, Marcu 2001). Это приводит к тому, что в случаях, когда возможны множественные интерпретации, но разметчики ограничены проставлением только одной метки, остается неясным, отражает ли разногласие между разметчиками неправильную интерпретацию (например, из-за отсутствия у разметчика необходимых знаний) или же неоднозначность интерпретации аннотируемого текста и, в некотором смысле, вынужденный выбор единственной метки из нескольких возможных.

Только в двух известных нам корпусах допускается аннотирование с использованием нескольких меток: Надкорпусная база данных коннекторов (НБДК), о которой речь пойдет ниже, и PDTB.

В корпусе PDTB 2.0 допускалось наличие более чем одного отношения между двумя фрагментами текста в трех случаях: *i*) отношение может быть выражено эксплицитным коннектором, который одновременно может выражать более одного отношения, например, *since* в (5) выше; *ii*) каждое из отношений между одинаковыми отрезками текста может быть выражено разными коннекторами, например, в многоэлементном коннекторе *but then* союз *but* выражает отношение «вопреки ожидаемому», а *then* – предшествование (6); *iii*) при отсутствии коннектора разметчики видят более одного отношения между фрагментами текста. В PDTB 3.0 к этим трем случаям был добавлен четвертый: *iv*) коннектор или другой показатель выражают одно дискурсивное отношение, но разметчик видит между фрагментами текста еще одно, имплицитное, отношение.

Каждому из четырех случаев соответствует своя разметка. Как видно из (5) в случае *i*) проставляется комбинированная метка с двумя выявленными отношениями. В случае *ii*) каждое из отношений размечается отдельно (6).

- (6) *Small businesses say a recent trend is like a dream come true: more-affordable rates for employee-health insurance, initially at least **But then** (PRECEDENCE) they wake up to a nightmare [wsj 0518]*
*Small businesses say a recent trend is like a dream come true: more-affordable rates for employee-health insurance, initially at least **But** (ARG2-AS-DENIER) **then they wake up to a nightmare** [wsj 0518]* (пример из Prasad et al. 2019: 53)

В случае *iii*), как и в случае *i*) проставляется комбинированная метка, а семантика имплицитного дискурсивного отношения определяется на основе возможной подстановки эксплицитного показателя (7):

- (7) *...there's also a price to pay* (IMPLICIT=REASON, ARG2-AS-DETAIL) **A diversified portfolio always under-performs an undiversified portfolio during those times when the investment in the undiversified portfolio is truly hot** [wsj 1453] (пример из Prasad et al. 2019: 53)

Наконец, в случае *iv*) каждое из отношений – имплицитное и эксплицитное – размечаются отдельно, как в *ii*), но с другим способом представления:

- (8) *We've got to get out of the Detroit mentality and* (EXPANSION.CONJUNCTION) Implicit=instead (EXPANSION.SUBSTITUTION.ARG2-AS-SUBST) **be part of the world mentality,**" declares Charles M. Jordan, GM's vice president for design... [wsj 0956]

Характерно, что, признавая важность для теоретических когнитивных исследований использования комбинированной метки в случае *i*) (Prasad et al. 2014: 933), разработчики корпуса как будто не замечают, насколько важно дать возможность исследователям изучать также потенциальную совместимость отношений в случае *ii*). Такую возможность разметка, принятая в PDTB, к сожалению, не предоставляет, и сводные таблицы дают только информацию по совмещению отношений в случае *i*) (см. Prasad et al. 2019).

3 Аннотации с неоднозначными коннекторами в НБДК

Надкорпусная база данных коннекторов (НБДК), созданная в ИПИ РАН, содержит тексты параллельных английского, немецкого, французского и итальянского подкорпусов НКРЯ. На 1.01.2025 в ней сформировано 25'547 двуязычных аннотаций. В них отражено не только дискурсивное отношение, выражаемое коннектором, но и такие параметры, как синтаксическая структура вводимого коннектором фрагмента текста, его позиция в нем, порядок следования фрагментов текста. Для полифункциональных языковых единиц (например, *точнее* может быть как сравнительной степенью прилагательного или наречия, так и коннектором) указана его функция в данном контексте (подробнее об устройстве и функциональных возможностях НБДК см. Инькова 2018).

В НБДК для интересующих нас случаев используется следующая разметка. Для случая *ii*), т.е. когда в составе многоэлементного коннектора можно выделить несколько 'самостоятельных', в аннотации проставляется необходимое количество отношений: их может быть два или более. В таблице 1 представлена разметка примера (4) выше с *a... тем временем*, выражающим два отношения. Заметим, что в итальянском переводе одноэлементный коннектор *intanto* переводит только отношение одновременности.

Контекст коннектора в оригинале	Признаки коннектора в оригинале	Контекст коннектора в переводе	Признаки коннектора в переводе
В руках он держал книгу. За ним следом – молодой дьяк. Заметив краем глаза, что справедливость вот-вот будет восстановлена, Виктор продолжил свой путь. А дьяк тем временем вырвал из рук мальчишки книгу, ударил этой книгой его несколько раз по голове и тоже поспешил обратно к церкви.	a тем временем сопоставительные одновременность повествовательное начальная p CNT q CNT	Teneva in mano un libro. Dietro a lui correva un giovane diacono con una barbetta rada e la faccia brufolosa. Notato con la coda dell'occhio che presto sarebbe stata fatta giustizia, Viktor continuò per la sua strada. Il diacono intanto strappò dalle mani del ragazzino il libro, con il quale lo colpì ripetutamente in testa, e si affrettò di nuovo a tornare in chiesa.	intanto одновременность повествовательное неначальная p CNT q CNT

Таблица 1. Двуязычная аннотация контекста коннектора, выражающего более одного значения, в НБДК

Такая же разметка возможна в принципе и для имплицитных отношений – случай *iii*) в PDTB –, но таких случаев в НБДК пока не зафиксировано, как и контекстов, соответствующих случаю *iv*). Можно было бы проставить в разметке итальянского перевода фрагмента текста с *a... тем временем* и имплицитное отношение сопоставления (помимо проставленного отношения

одновременности), но поскольку в НБДК фиксируются реальные употребления коннекторов (в данном примере ит. *intanto*), то разметка фиксирует характеристики именно переводного эквивалента, а не контекста, в котором он употреблен. Хотя и эта информация важна, в частности, для изучения того, какие отношения могут быть имплицированы.

Для случая *i*), т.е. когда разметчик затрудняется выбрать одно из выражаемых полисемичным коннектором отношений, предусмотрена комбинированная метка, сочетающая несколько отношений. Так, для *точнее* в аннотации, представленной в таблице 2, как и для его переводного эквивалента, была проставлена метка «коррекция/спецификация». Фрагмент текста, вводимый коннектором, может быть понят как переход от общего (Достоевский) к частному (его идеология), т.е. как спецификация, или как своего рода поправка-уточнение, т.е. как коррекция. Контекст не позволяет разметчику выбрать то отношение, которое однозначно соответствовало бы коммуникативному заданию говорящего.

Контекст коннектора в оригинале	Признаки коннектора в оригинале	Контекст коннектора в переводе	Признаки коннектора в переводе
Такой образ в стиле самого Достоевского, точнее , его идеологии, между тем как образ единого духа глубоко чужд ему.	Точнее коррекция / спецификация без предикации начальная p CNT q CNT	Quest'immagine è nello stile di Dostoevskij, o più esattamente della sua ideologia, mentre l'immagine di un unico spirito gli è affatto estranea.	O più esattamente коррекция / спецификация без предикации начальная p CNT q CNT

Таблица 2. Двухязычная аннотация контекста коннектора с комбинированной меткой в НБДК

Наконец, если разметчик затрудняется в определении отношения, например, в том случае, если семантика коннектора еще не была точно определена, то он может проставить метку TBD (To Be Defined). Эта метка сама по себе не представляет интереса, но позволяет разметчикам создавать аннотации с еще неизученными коннекторами.

По всем трем типам меток в НБДК можно получить количественные данные. Например, для русских коннекторов было сформировано 1669 аннотаций, где проставлено два и более отношений. Это составляет 6,54%. Комбинированных меток гораздо меньше – 78, т.е. 0,3%. Для сравнения: в PDTB для эксплицитных отношений такие метки были проставлены в 999 из 18 459 аннотаций, т.е. в 5,4%, но это различие связано с разной методикой разметки. Количество аннотаций с меткой TBD постоянно меняется (на 1.01.2025 их было 2660).

4 О чем говорят аннотации с несколькими метками?

В цитированных работах, посвященных использованию аннотирования с несколькими метками, подчеркивается важность исследования этих меток. В частности, ставятся следующие вопросы:

- если коннектор размечен как выражающий несколько отношений, то все ли размеченные комбинации отношений являются допустимыми или некоторые из них можно признать аномальными?
- выводимы ли какие-то отношения из контекста, а не из значения коннектора?
- какие отношения дистрибутивно схожи, т.е. встречаются в одном контексте?

Чтобы ответить на эти и другие вопросы, необходимо, как нам кажется, разграничить два случая. В первом из них мы имеем дело с сочетаниями отношений, выражаемыми многоэлементными коннекторами, условно случай *ii*) в модели PDTB. Как правило, это сочетания сочинительного союза с «конкретизатором» (в терминологии русской грамматики) или с коннектором наречной природы, как в (4) или (6) выше. С одной стороны, эти сочетания позволяют нам выявить отношения, которые могут употребляться в одном контексте, а с другой – сочетаемость коннекторов обусловлена не только концептуальной совместимостью дискурсивных отношений, которые они выражают, но и их синтаксическими свойствами. Например, сочинительные или подчинительные союзы не могут сочетаться между собой. Кроме того, если показатели некоторых отношений избирательны в своих сочетаниях, что позволяет выявить некоторые закономерности (например, *хотя* в составе двухкомпонентного коннектора может сочетаться только с союзом *но*,

впрочем сочетается только с союзами *а* или *но*), то показатели других отношений не знают таких ограничений. Так, показатель отношения спецификации *например* сочетается со всеми сочинительными союзами: *и, а, но, или* и даже *то есть* (9).

- (9) Осторожно предполагая, что новоискровцы – по сути оппортунисты и лакеи буржуазии, он отмечает, что пока это проявляется как «оппортунизм в организационных вопросах»; что ему можно противопоставить – например, право ЦК распускать местные организации (*то есть, например, Заграничную лигу: централизм так централизм*). [Л. А. Данилкин. Ленин: Пантократор солнечных пылинок (2017)]

Наконец, сочетание с сочинительным союзом может объясняться не столько желанием говорящего выразить еще какое-то дискурсивное отношение, сколько синтаксическими причинами. Например, коннектор *в то же время* не может связывать два однородных сказуемых, не прибегая к помощи союза *и*, и в НБДК сочетание соединительных отношений и отношений одновременности занимает первую позицию как среди аннотаций с несколькими отношениями (157 из 1669, т.е. 9,4%), так и среди аннотаций, сформированных для отношения одновременности (157 из 707, т.е. 22,2%).

Второй случай – комбинированная метка, сочетающая значения полисемичного коннектора, и здесь необходимо прояснить, что именно мы размечаем: отношение, т.е. то, как мы понимаем связь между двумя ситуациями, или коннектор? Получается некий замкнутый круг. В PDTB отношение проставляется на основе семантики коннектора, но разметчик может «чувствовать», что коннектор выражает еще какое-то отношение. Это происходит, как правило, под влиянием контекста. Например, союз *и* всегда сохраняет соединительное значение, которое является для него базовым. Другое дело, что под влиянием смыслового соотношения фрагментов текста, у него можно увидеть значение временного следования, результативно-следственное и другие, которые выводятся, в том числе, с учетом наших экстралингвистических знаний, но которые, тем не менее, фиксируются словарями. Среди комбинированных меток PDTB вторая по частотности (138 из 999 аннотаций) как раз приходится на союз *and*, «совмещающий» причинное и соединительное значения (Prasad et al. 2014: 933).

В НБДК было принято решение размечать союз *и* всегда только как соединительный. Но здесь интересно посмотреть, с каким семантическим соотношением фрагментов текста совместимо это соединительное значение. Это значит, однако, что мы будем фиксировать не столько отношение или значение коннектора, сколько семантические свойства контекста, в котором он употреблен. Так, для показателя отношения переформулирования в итальянском языке *cioè* ‘то есть’ было показано, что при всем разнообразии семантических отношений между фрагментами текста, с которым он совместим, он не может быть употреблен, если между ними есть причинное отношение (Manzotti 1999). Это важный параметр, позволяющий более четко описать дискурсивные отношения. Неправомерно, на наш взгляд, приписывать коннектору то, что мы выводим из контекста. Например, *в то же время* может употребляться для связи фрагментов текста, находящихся в отношении антонимии. Это способствует его интерпретации как сопоставительного (или даже противительного), но при этом его базовое значение одновременности сохраняется, как в (10):

- (10) Крымов, посмеиваясь *и в то же время* серьезно, перебил: – А понадобится, и вас, Греков, с большевистской кашей съедят. [В. С. Гроссман. Жизнь и судьба (1959)]

Можно ли здесь говорить о «совмещении» значений коннектора, если отношение противопоставления диктуется контекстом? И наоборот, в некоторых контекстах сопоставительное значение полностью вытесняет временное (11), следовательно, здесь также несправедливо говорить о каком-либо совмещении значений.

- (11) По нашим данным, 70% девушек твердо собираются покинуть свой город, и 17% пока колеблются; *в то же время* среди юношей уверенных – 54%, а колеблющихся – 28%. [Ю. Ф. Флоринская, Т. Г. Рощина. Жизненные планы выпускников школ из малых городов (2004) // «Человек», 2005]

Контексты, в которых коннекторы действительно неоднозначны, т.е. те, что не позволяют выбрать то или иное значение полисемичного коннектора, крайне редки; ср. приведенный в таблице 2 пример для *точнее*. Неслучайно поэтому, что в НБДК проставлено такое малое количество комбинированных меток (0,3%).

В комбинированных метках также, как и в случае аннотаций с несколькими метками отношений, интересно проанализировать сочетание отношений, между которыми разметчик затрудняется сделать выбор. Даже самый предварительный анализ позволяет говорить о том, что разметчик затрудняется выбрать между семантически близкими отношениями. В НБДК одной из распространенных комбинированных меток является метка «Одновременность / Условие» для союзов *когда* и *как*. В PDTB эта метка (Contingency.Condition / Temporal.Synchrony) для коннектора *when* тоже является одной из наиболее частотных (Prasad et al. 2014: 933). Сочетания отношений в комбинированных метках позволяет, таким образом, дополнить анализ степени семантической близости отношений, проводимый на основе их структурированных определений (подробнее см. Инькова, Кружков 2023a,b).

5 Заключение

В заключение сказанного хотелось бы сформулировать дальнейшие этапы исследования дискурсивных отношений. На наш взгляд, следующим важным этапом в исследовании дискурсивных отношений должно стать аннотирование семантических свойств фрагментов текста, между которыми установлено дискурсивное отношение. Это позволит определить, какие отношения могут встречаться в одном контексте, т.е. дистрибутивно близки, а также предсказать, каким образом может быть разрешена неоднозначность; ср. англ. *so* в (1) и (2) выше. Это, в свою очередь, поможет установить, какие сочетания отношений являются нормальными, а какие «аномальными», что повысит качество разметки. При этом особенности контекста должны быть сформулированы на доступном машине языком, чтобы улучшить автоматическую обработку текста, которая особенно затруднена в случае имплицитных дискурсивных отношений. Они смогут, таким образом, определяться не путем подстановки соответствующего по смыслу коннектора, которые, как мы видели, неоднозначны, а на основе семантических свойств контекста.

Благодарности

Работа выполнена в Федеральном исследовательском центре «Информатика и управление» Российской академии наук с использованием ЦКП «Информатика» ФИЦ ИУ РАН.

Acknowledgements

The research was carried out using the infrastructure of the Shared Research Facilities “High Performance Computing and Big Data” (СКР “Informatics”) of FRC CSC RAS (Moscow).

References

- [1] Aroyo L., Welty Ch. Truth is a lie: Crowd truth and the seven myths of human annotation // AI Magazine. – 2015. – Vol. 36:1. P. 15-24.
- [2] Basile V., Fell M., Fornaciari T., Hovy D., Paun S., Plank B., Poesio M., Uma A. We need to consider disagreement in evaluation // Proceedings of the 1st Workshop on Benchmarking: Past, Present and Future. – Bangkok, Thailand, 2021. P. 15-21. Access mode: <https://aclanthology.org/2021.bppf-1.pdf>.
- [3] Carlson L., Marcu D. Discourse Tagging Reference Manual. ISI-TR-545. University of Southern California Information Sciences Institute, 2001. Access mode: <http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>.
- [4] Inkova O. Linguospecificity of connectives: methods and parameters of description [Lingvospecificnost' konnektorov: metody i parametry opisaniya] // O. Inkova (ed.), Semantics of connectives: a contrastive study [Semantika konnektorov: kontrastivnoe issledovanie]. – Moscow: TORUS PRESS, 2018. P. 5-23.
- [5] Inkova O., Kruzikov M. Evaluating the degree of discourse relations semantic affinity: methods and instruments [Stepen semanticheskoy blizosti diskursivnykh otnoshenij: metody i instrumenty rascheta] // Informatics and applications. – 2023b. – Vol. 17:4. P. 46-53.

- [6] Kobozeva I.M., Serdobolskaya N.V., Kryukova A.I., Pilyugina D.A. Creating a database of Modern Russian clause linkers: principles, problems, results [Sozdanie bazy dannyx konnektorov sovremennogo russkogo yazyka: principy, problemy, rezultaty] // Tyumen State University Herald. Humanities Research. Humanities. – 2023. – Vol. 9:4. P. 36-47. Access mode: <https://doi.org/10.21684/2411-197X-2023-9-4-36-47>
- [7] Long W., Siddharth N., Webber B. Multi-Label Classification for Implicit Discourse Relation Recognition // Computing Research Repository. – 2024. – Vol. arXiv:2406.04461 [cs.CL]. Access mode: <https://arxiv.org/abs/2406.04461>.
- [8] Lopez Cortez S. M., Jacobs C. L. Incorporating Annotator Uncertainty into Representations of Discourse Relations // Computing Research Repository. – 2023. – Vol. arXiv:2308.07179 [cs.CL]. Access mode: <https://arxiv.org/abs/2308.07179>.
- [9] Mann W.C., Matthiessen Ch.M.I.M., Thompson S. A. Rhetorical structure theory and text analysis. ISI Research Report ISI/RR-89-242, 1989. DOI: 10.1075/pbns.16.04man.
- [10] Manzotti E. Spiegazione, riformulazione, correzione, alternativa: sulla semantica di alcuni tipi e segnali di parafrasi // L. Lumbelli, B. Mortara Garavelli (a c. di), Parafrasi. Dalla ricerca linguistica alla ricerca psicopedagogica. – Alessandria: Edizioni dell’Orso, 1999. P. 169-206.
- [11] Marchal M., Scholman M., Yung F., Demberg V. Establishing annotation quality in multi-label annotations // Proceedings of the 29th International Conference on Computational Linguistics. – Gyeongju, Republic of Korea, 2022. P. 3659-3668. Access mode: <https://aclanthology.org/2022.coling-1.322/>.
- [12] Nuriev V., Buntman N., Inkova O. Machine translation of Russian connectives into French: errors and quality failures // Informatics and applications. – 2018. – Vol. 12:2. P. 105-113.
- [13] PDTB Research Groupe. The Penn Discourse Treebank 2.0 Annotation Manual. Technical Report IRCS-08-01. Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania, 2008. Access mode: <https://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>.
- [14] Prasad R., Webber B., Joshi A. Reflections on the Penn Discourse TreeBank, Comparable Corpora, and Complementary Annotation // Computational Linguistics. – 2014. – Vol. 40:4. P. 921-950. Access mode: <https://aclanthology.org/J14-4007.pdf>.
- [15] Prasad R., Webber B., Lee A., Joshi A. Penn Discourse Treebank Version 3.0 // Linguistic Data Consortium. – 2019. – Vol. LDC2019T05. Access mode: <https://catalog.ldc.upenn.edu/LDC2019T05>.
- [16] Scholman M., Dong T., Yung F., Demberg V. DiscoGeM: A crowdsourced corpus of genre-mixed implicit discourse relations // Proceedings of the Thirteenth Language Resources and Evaluation Conference. – Marseille, France, 2022. P. 3281-3290. Access mode: <https://aclanthology.org/2022.lrec-1.351/>.
- [17] Spooren W., Degand L. Coding coherence relations: Reliability and validity // Corpus Linguistics and Linguistic Theory. – 2010. – Vol. 6:2. P. 241-266.
- [18] Stede M. RST revisited: Disentangling nuclearity // C. Fabricius-Hansen & W. Ramm, editors, Subordination’ versus ‘coordination’ in sentence and text from a cross-linguistic perspective. John Benjamins: Amsterdam, 2008. P. 33-58.
- [19] The Russian National Corpus [Nacional’nyi korpus russkogo yazyka]. Access mode: <https://ruscorpora.ru/>
- [20] Webber B. What excludes an alternative in coherence relations? // Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013). – Potsdam, Germany, 2013. – P. 276-287.
- [21] Zufferey S., Degand L. Connectives and Discourse. – Cambridge: Cambridge University Press, 2024.