

Defining discourse relations: Supracorpora database of connectives

Inkova O.

Institute of Informatics Problems, FRC

CSC RAS / Moscow, Russia

University of Geneva, Geneva,

Switzerland

Olga.Inkova@unige.ch

Abstract

The research is focused on definitions of discourse relations, a topic that is currently little-studied. The paper gives a brief overview of existing solutions for discourse relations definitions: Rhetorical Structure Theory (RST), Segmented Discourse Representation Theory (SDRT), Penn Discourse Treebank (PDTB), and Cognitive approach to Coherence Relations. The author shows criteria used to define a discourse relation, or, in case of a narrower definition, a logical-semantic relation, in these approaches and outlines the shortcomings of the described definitions. The author also describes the principles used to build the classification and the definitions of logical-semantic relations (LSR) in the Supracorpora Database of connectives (SDB). The classification is based on four basic semantic operations upon which rests every LSR's definition: implication, location on the chronological scale, comparison, correlation between specific and general or an element and a set. The classification consistently distinguishes the levels at which the LSR can be established: propositional, illocutionary, and metalinguistic. Each LSR is defined on the basis of these two criteria. Thus, for example, for the LSR of alternative based on the comparison operation, one has the choice between the LSR of propositional, illocutionary and metalinguistic alternative (*We will go to the mountains or to the sea vs. Put the gun away, or are you scared? vs. The symbol of the year or, simply speaking, cutie-pie*). In case of LSRs based on implication or comparison, the polarity criterion is added, distinguishing whether the LSR is established between p and q or their negative correlates $\neg p$ and $\neg q$ are also to be taken into account in order to obtain a correct interpretation (cf. well-known descriptions of how the Russian conjunction *no* 'but' functions). In addition, semantic and pragmatic characteristics of the context are also considered in the classification. For example, in the case of the LSR of specification and generalization, the semantic correlation between p and q (together with their intensional and extensional interpretations) is taken heed of. Several definitions of LSR and corresponding examples are provided. Thus, the LSR of extensional specification is defined as follows: based on the operation of correlation between the general and the particular; established at the propositional level; X contains a generalized notion or state of things p ; Y contains a more particular q -notion, limiting p -extensional. And the LSR of intensional specification is defined as follows: based on the operation of correlation between the general and the particular; established at the metalinguistic level; X contains a generalized concept or state of things p ; Y contains a more particular q -notion, limiting p -intensional. The definitions used in the SDB definitions make it possible to evaluate, on the basis of the proposed criteria, the semantic closeness of relations and increase the level of consistency in the work of experts and annotators. That in turn increases the value of the annotated material, and therefore its reliability.

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

DOI: 10.28995/2075-7182-2021-20-328-338

Определения дискурсивных отношений: опыт Надкорпусной базы данных коннекторов

Инькова О.Ю.

ИПИ ФИЦ ИУ РАН / Москва, Россия

Женевский университет / Женева,

Швейцария

Olga.Inkova@unige.ch

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

1 Вводные замечания

Настоящее исследование посвящено мало изученной проблеме определения дискурсивных отношений. В работах, им посвященным, рассматриваются, как правило, лишь вопросы их классификации, также вызывающие многочисленные дискуссии; см., например, [Roze 2013; Chircos 2014; Benamara, Taboada 2015; Bunt, Prasad 2016; Demberg et al. 2019, Гончаров 2021]. Хотя эти два вопроса тесно связаны, вопрос об определениях дискурсивных отношений и критериях, лежащих в их основе, заслуживает отдельного рассмотрения.

Как известно, существуют два подхода к определению отношений:

- отношение определяется на основе семантики его прототипического показателя; см., [Halliday 1985; Knott 1996; Stede, Umbach 1998; Alonso et al. 2002; Fraser 2009] и др.;
- отношение определяется на основе коммуникативного задания говорящего или семантических и прагматических характеристик контекста; см., например, [Grosz, Sidner 1986; Mann, Thompson 1988; Asher, Lascarides 2003; Prasad et al. 2017]. Тем не менее, даже в этом случае во многих корпусах предлагается список прототипических показателей, которые в процессе аннотирования помогают установить, о каком отношении идет речь; см. ANNODIS [Muller et al. 2012], [PDTB Project].

Мы начнем с краткого обзора существующих решений для определения дискурсивных отношений, а именно трех подходов, используемых для аннотирования текстов (Теория риторической структуры, Теория сегментной репрезентации дискурса, Пенсильванский корпус), а также когнитивного подхода к отношениям связности. Затем представим принципы, разработанные для определения отношений и их аннотирования в Надкорпусной базе данных коннекторов¹, и приведем несколько примеров таких определений.

2 Теория риторической структуры

В Теории риторической структуры (RST) отношение связности определяется на основе коммуникативных намерений и представлений говорящего (Г), а также представлений слушающего (С) по четырем параметрам: ограничения на сателлит (Ст); ограничения на ядро (Я); ограничения на сочетание ядра и сателлита; достигаемый эффект. Приведем в качестве примера определение отношения Evidence² и иллюстрирующий его пример, заимствованный с сайта RST [<http://www.sfu.ca/rst/>]:

Отношение: Evidence.
Ограничения на Я: С, по мнению Г, не верит в Я в достаточной степени.
Ограничения на Ст: Г верит в Ст или считает его достаточно достоверным.
Ограничения на сочетание Я+С: Вера С в Ст увеличивает его веру в Я.
Эффект: Вера С в Я увеличена.

- (1) 1) Darwin as a Geologist 2) he tends to be viewed now as a biologist, 3) but in his five years on the Beagle his main work was geology, 4) and he saw himself as a geologist. 5) His work contributed significantly to the field. 6) Scientific American, Sandra Herbert, May 1986, o. 116.

Согласно предлагаемому на сайте описанию, в (1) отношением Evidence связаны между собой дискурсивная единица 1), которая является заголовком статьи, и дискурсивные единицы 3)-5). Единицы 1) и 2) связаны отношением Concession 'Уступка'. Какими отношениями связаны между собой 2) и 3), а также 3), 4) и 5) остается из анализа неясным. Неясной остается также функция дискурсивной единицы 6). Она просто оставлена на верхнем уровне схемы.

¹ Представительный фрагмент Надкорпусной базы доступен по ссылке:

<http://a179.frcsc.ru/PublicLingvoProjects/main.aspx>.

² Мы предпочитаем сохранять английские названия отношений, чтобы не создавать дополнительную терминологическую путаницу, поскольку как названия отношений, так и объем даже одинаковых терминов не совпадают в разных классификациях.

Определения RST можно упрекнуть в некоторой тавтологичности. Особенно это касается многоядерных отношений. Так, отношение Disjunction определено следующим образом: а) элемент представляет собой (не обязательно исключаящую) альтернативу другому; б) С распознает, что связанные элементы являются альтернативами. Отношение Sequence ‘Следование’: а) между ситуациями, описанными в Я, существует следование во времени; б) С распознает следование во времени между Я [http://www.sfu.ca/rst/]. Однако на каком основании – семантическом, прагматическом, ... – слушающий распознает эти отношения остается неясным. Эта тавтологичность отчасти преодолена в пособии по аннотированию RST Treebank. Из определения исчезает семантический эффект, и Sequence, например, определяется так: “A Sequence is a multinuclear list of events presented in chronological order” [Carlson, Marcu 2001: 67].

Кроме того, в RST набор отношений разделен на две группы, соответствующие *grosso modo* семантическим (Subject Matter Relations) и прагматическим (Presentational Relations) отношениям, внутри которых они даны списком, что не позволяет увидеть, что некоторые отношения, в том числе находящиеся в разных группах, имеют общие семантические свойства.

3 Теория сегментной репрезентации дискурса (SDRT)

В отличие от RST, SDRT ставит своей целью моделировать процесс интерпретации текста, определяя правила выведения дискурсивных отношений, которые разделены на три группы: действующие на пропозициональном уровне, или семантические (content-level relations), действующие на уровне высказывания, или прагматические (meta-talk relations), и структурные (text-structuring relations), названные так, поскольку они накладывают определенные структурные ограничения на связываемые ими дискурсивные единицы [Asher, Lascarides 2003]³. В рамках этих групп семантическая близость входящих в них отношений не эксплицируется.

Само отношение определяется на основе семантической, прагматической и лексической информации при помощи порождающих правил. Их формализация менялась в ходе развития SDRT, но основные ее положения, сформулированные, начиная с [Asher 1996], на языке связующей логики (Glue Logic), сводятся к следующему. Отношения считаются трехместными предикатами $R(\alpha, \beta, \lambda)$, указывающими на то, что α и β (пропозиции в сочетании с характеризующими их коммуникативными намерениями) связаны отношением R в дискурсивном сегменте λ . $?(\alpha, \beta, \lambda)$ означает, что между α и β существует дискурсивное отношение, но мы еще не знаем, какое именно. Язык связующей логики включает также формулы типа [A]K, указывающие, что A является элементом пропозиционального содержания минимальной дискурсивной единицы K, а также классический оператор импликации \rightarrow и условный немонотонный оператор $>$.

Рассмотрим, как представлено отношение Explanation, основанное, как и Evidence в RST, на причинных отношениях, на примере, приводящемся практически во всех работах, посвященных теоретическим принципам SDRT.

(2) Max fell. John pushed him. [пример из Asher 1993]

Если α и β связаны между собой и K_α описывает положение вещей, в котором некоторый субъект падает, а K_β – положение вещей, где того же субъекта толкают, мы можем предположить, на основании наших знаний о мире, что второе положение вещей может быть причиной первого, что выражается предикатом Cause_D (β, α, λ). Нижний индекс D в предикате Cause означает, что мы имеем дело с Дискурсивными указаниями на причинность. Формальное представление (2) дано в (3):

(3) $(?(\alpha, \beta, \lambda) \wedge [\text{Fall}(e_1, y)]K_\alpha \wedge [\text{Push}(e_2, x, y)]K_\beta) \rightarrow \text{Cause}_D(\beta, \alpha, \lambda)$

Затем дискурсивное отношение характеризуется его семантическим эффектом. Семантический эффект отношения Explanation – установление причинных связей между положениями вещей K_α и K_β , из чего вытекает также, что K_α непосредственно следует за K_β .

³ На самом деле групп отношений семь, но четыре из них касаются особенностей диалогического текста.

$$(4) \quad \text{Explanation}(\alpha, \beta) \Rightarrow (\text{cause}(e_\beta, e_\alpha) \\ ((\text{cause}(e_\beta, e_\alpha) \wedge \text{event}(e_\alpha)) \Rightarrow e_\beta < e_\alpha$$

Таким образом, определение состоит из двух частей: семантических и прагматических характеристик контекста и семантического эффекта отношения. То же можно сказать и про структурные виды отношений, к которым принадлежит Contrast. К ограничениям, которые оно накладывает на контекст, относятся изоморфизм синтаксической структуры и контрастная тема (которая может быть выражена, например, предикатами противоположной полярности или семантики; см. [Asher et Lascarides 2003]):

$$(5) \quad (?(\alpha, \beta, \lambda) \wedge \text{Structurally_similar}(\alpha, \beta) \wedge \text{Contrasting_themes}(\alpha, \beta)) > \text{Contrast}(\alpha, \beta, \lambda)$$

Что касается семантического эффекта отношения Contrast, то из него следует, что α и β имеют несовместимые импликатуры, т. е. пропозициональное содержание α позволяет вывести факт, отрицание которого может быть выведено из пропозиционального содержания β [Bras 2008: 47].

Очевидно, однако, что формальный аппарат SDRT слишком сложен, чтобы использовать его при аннотировании корпусов, а набор отношений – слишком невелик: их всего 12. В работе [Reese et al. 2007: 8], которая является по сути пособием по аннотированию, к ним добавлены еще два, а формализация, особенно пропозиций, исчезает. Отношения разделяются на две группы, сочиняющие и подчиняющие, и определяются описательно. Ср.: “Explanation(α, β) holds when the main eventuality of β is understood as the cause of the eventuality in α . Explanation has temporal consequences, viz. that the eventuality described in β precedes (or overlaps) the eventuality described by α ” [Reese et al. 2007: 12]. Ср. определения в аннотированном корпусе французских текстов ANNODIS [Muller et al. 2012], также использующем SDRT.

4 Пенсильванский корпус (PDTB)

PDTB дает классификацию более узкой группы дискурсивных отношений, а именно тех, которые потенциально могут быть выражены коннекторами, т.е. логико-семантических отношений (ЛСО). Они сгруппированы по степени семантической близости и разделены на четыре группы: Temporal, Contingency, Comparison, Expansion, в каждой из которых выделяются подтипы. Так, в рамках временных ЛСО выделяются две группы: синхронные и асинхронные, а в рамках последней следование и предшествование. В рамках групп Contingency и Comparison дальнейшая классификация различает ЛСО, действующие на пропозициональном уровне (Cause, Condition, Concession, Contrast) и действующие на уровне речевого акта (Pragmatic Cause, Pragmatic Condition, Pragmatic Contrast). Однако только первая группа задана на основе четкого семантического критерия, распространяющегося на всю совокупность входящих в ее состав ЛСО и позволяющего отличить эту группу ЛСО от других. Определения, данные для остальных групп, не позволяют этого сделать. Так, Contingency определяется как отношение, при котором “one argument provides the reason, explanation or justification for the situation described by other” [PDTB Research Group 2019: 19]. Под это определение явно не попадает подгруппа отношений Condition, включающая отношения, при которых “one argument presents a situation as unrealized (the antecedent), which (when realized) would lead to the situation described by the other argument (the consequent)” (с. 22).

В группу Comparison, основанную на сходстве или несходстве соединяемых ситуаций, включено, напротив, наряду с Contrast и Similarity, также Concession, в семантику которого принято включать, причинный компонент, а именно отрицание ожидаемой причинной связи между ситуациями. Разработчики PDTB (с. 24) даже подчеркивают, что именно этот компонент отличает Concession от Contrast, где этот компонент отсутствует. Место Concession – скорее в группе Contingency, где оно логично соседствовало бы с ЛСО Negative Condition, также задействующим отрицание имплицативной связи между ситуациями.

Наибольшая непоследовательность наблюдается, однако, в группе Expansion, объединяющей отношения, “that expand the discourse and move its narrative or exposition forward”⁴; см. рис. 1.

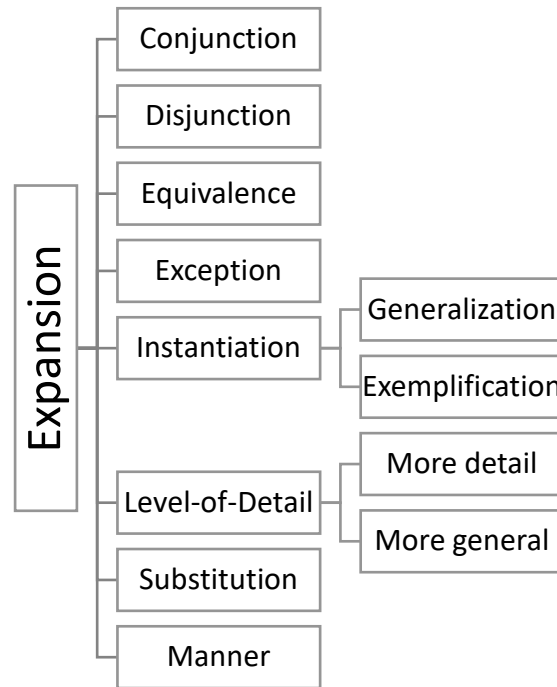


Рисунок 1. Группа ЛСО Expansion в PDTB [PDTB Research Group 2019].

Часть из отношений группы Expansion, как мы покажем в разделе 6, логично нашли бы свое место в группе Comparison: Conjunction, Disjunction, Substitution, Equivalence и подвид More general ЛСО Level-of-Detail. ЛСО Exception, подвид More detail ЛСО Level-of-Detail и обе разновидности Instantiation (Generalization, Exemplification) могут быть объединены в группу ЛСО, выражающих отношения элемента и множества или общего и частного. Кроме того, некоторые из этих ЛСО (например, Disjunction) могут быть установлены как на пропозициональном уровне, так и на уровне речевого акта, однако в классификации PDTB это никак не оговаривается, в отличие от ЛСО групп Contingency и Comparison, а Equivalence и подвид More general ЛСО Instantiation устанавливаются на метаязыковом уровне, так как касаются выбора языковых средств описания одной и той же ситуации, как и метаязыковая альтернатива и замещение по адекватности описания. Этот уровень не выделяется в PDTB. Таким образом, в группе Expansion остается только ЛСО Manner, добавленное, заметим, только в последней версии классификации (3.0).

5 Когнитивный подход к отношениям связности (CCR)

Когнитивный подход к отношениям связности, разработанный в работах Т. Сандерса (см., из последних [Sanders et al. 2018]), определяет отношения связности, которые являются логико-семантическими, через набор «когнитивно базовых» признаков:

- полярность (отрицательная vs. положительная), т. е. необходимо ли для интерпретации отношения привлекать отрицание одного из соединяемых положений вещей (к отрицательным относятся, например, уступительные отношения);
- базовая операция (аддитивная vs. причинная); причинной считается операция, основанная на импликации, соответственно, в группу причинных входят условные, причинные, целевые, следственные и уступительные отношения;

⁴ Отметим в этой связи, что, например, в SDRT или в Лингвистической модели дискурса Л. Поланьи термин Narration объединяет временные отношения как раз на том основании, что они позволяют двигаться вперед сюжетной канве; ср. характеристику Narratives «giving a next episode of a story» в [Polanyi 1995: 17].

- источник связности (объективный vs. субъективный), иначе говоря, насколько говорящий «вовлечен» в устанавливаемое отношение; эта оппозиция соответствует оппозиции между семантическими (объективными) и прагматическими (субъективными) отношениями в других классификациях (см., например, выше PDTB);
- порядок сегментов (базовый vs. небазовый); этот признак применяется только к группе отношений, основанных на причинной операции, и позволяет различать следствие (с базовым порядком, где причина как хронологически, так и в линейной структуре высказывания, предшествует следствию) и причину (с небазовым порядком, где в линейной последовательности высказывания следствие предшествует причине).

Этих признаков, однако, явно недостаточно, чтобы различать отношения. Не спасает дела и введение дополнительных признаков (среди которых, например, спецификация, альтернатива, условность). Так, обе разновидности отношений Instantiation и Level-of-Detail⁵ имеют одинаковый набор не только основных признаков (положительная полярность, аддитивная операция, объективный/субъективный источник связности, признак порядка не применяется), но и дополнительный (specificity ‘специфичность’) [Sanders 2018: 61].

6 Определение ЛСО в Надкорпусной базе данных коннекторов

В Надкорпусной базе данных коннекторов (НБД) используется классификация, в основу которой положены четыре базовые семантические операции, или механизма, на которые опирается то или иное ЛСО: импликация, расположение на шкале времени, сравнение, соотнесение частного и общего или элемента и множества. Классификация последовательно различает уровни, на которых может быть установлено ЛСО: пропозициональный уровень, уровень высказывания (иллокутивный), метаязыковой; подробнее см. [Инькова 2019]. Соответственно, каждое ЛСО может определяться на основе этих двух критериев, к которым добавляется еще один, характеризующий ЛСО, основанные на импликации и сравнения: полярность, т.е. устанавливается ли ЛСО непосредственно между p и q или же при его интерпретации должны быть учтены также их отрицательные корреляты $\neg p$ и $\neg q$. Кроме того, учитываются семантические и прагматические характеристики контекста.

На основе этих параметров впоследствии были сформулированы структурированные определения ЛСО. В настоящем исследовании мы приводим несколько таких определений и примеры, их иллюстрирующие.

Соединительное ЛСО (6)

- операция сравнения, устанавливающая сходство p и q относительно некоторого «общего знаменателя»;
 - пропозициональный уровень;
 - p и q – положения вещей, не связанные никаким другим ЛСО⁶.
-

- (6) Но эта многопланность и полифоничность мистерии чисто формальная, и самое построенное мистерии не позволяет содержательно развернуться множественности сознаний с их мирами. [М. М. Бахтин. Проблемы поэтики Достоевского (1963)]

⁵ В версии 2.0 [PDTB Research Group 2008], которая использована в [Sanders et al. 2018], ЛСО сгруппированы несколько по-другому, чем в версии 3.0, и подвиду More detail ЛСО Level-of-Detail соответствует Specification, а подвид More general рассматривался вместе с Generalization.

⁶ Т.н. «и-отношение» накладывает минимальные ограничения структурного и семантического характера и определяется, как правило, отрицательно. Единственное ограничение – наличие «общего знаменателя» (ср. Commun integrator, или Gemeinsame Einordungsinstanz, в [Lang 1977]).

Перифрастическое переформулирование ⁷ (7)	Обобщающее переформулирование ⁸ (8)
<ul style="list-style-type: none"> • операция сравнения, устанавливающая сходство p и q; • метаязыковой уровень; • p и q – описания одного и того же положения вещей; • p и q имеют одинаковый интенционал. 	<ul style="list-style-type: none"> • операция сравнения, устанавливающая сходство p и q; • метаязыковой уровень; • p и q – описания одного и того же положения вещей; • q имеет более бедный интенционал (т.е. является более общим описанием).

- (7) Греция неплатежеспособна, следовательно, возникает прагматический вопрос: как сократить нарастающий госдолг и повернуть вспять тенденцию сокращения объема производства? *Другими словами*: как увеличить выпуск и сократить долг? [Г. Колодко. Как избежать африканизации Греции // «Эксперт», 2015]
- (8) он часто брал меня на переговоры в качестве переводчика, и тут уж приходилось соответствовать и тону беседы, и её скорости, и свободе неадекватной фразеологии (особенно со стороны отечественного партнёра) и некоторой специфики самого разговора. *Короче говоря*, приходилось выкручиваться, и делала я это весьма ловко. [Невеста // «Туризм и образование», 2000.06.15]

Пропозициональная альтернатива (9)	Иллокутивная альтернатива ⁹ (10)	Метаязыковая альтернатива ¹⁰ (11)
<ul style="list-style-type: none"> • операция сравнения, устанавливающая несходство p и q; • пропозициональный уровень; • p и q – положения вещей, имеющие статус гипотезы; • говорящий предлагает сделать выбор между p и q. 	<ul style="list-style-type: none"> • операция сравнения, устанавливающая несходство p и q; • уровень высказывания; • P – речевой акт, q – положение вещей, имеющее статус гипотезы и ставящее под сомнение обоснованность P; • говорящий предлагает сделать выбор между P и q. 	<ul style="list-style-type: none"> • операция сравнения, устанавливающая несходство p и q; • метаязыковой уровень; • p и q – возможные описания одного и того же положения вещей r; • говорящий предлагает сделать выбор между p и q.

- (9) Квартирохозяйку тоже можно заинтересовать. *Или* запугать. [А. и Г. Вайнеры. Эра милосердия (1975)]
- (10) – Моя мамаша в Петракове. Гуслинский даже сел. – А как же... а эта дама? Мы же ее называли мамашей, *или* я преспокойно сошел с ума! А? [Н. А. Тэффи. Тонкая психология (1911)]
- (11) Она есть его тень *или*, по Гегелю, диалектический антитезис. [С. Булгаков. У стен Херсониса (1922)]

⁷ Equivalence в PDTB.

⁸ Подвид More general LCO Level-of-Detail в PDTB.

⁹ Отсутствует в PDTB.

¹⁰ Отсутствует в PDTB.

Пропозициональное замещение (12)	Замещение по дескриптивной адекватности ¹¹ (13)
<ul style="list-style-type: none"> • операция сравнения, устанавливающая несходство p и q; • пропозициональный уровень; • p – положение вещей, осуществление которого можно было бы ожидать; • q – положение вещей, не соответствующее ожиданиям; • p отвергается, принимается q. 	<ul style="list-style-type: none"> • операция сравнения, устанавливающая несходство p и q; • метаязыковой уровень; • p – описание положения вещей r, которое можно было бы ожидать; • q – описание того же положения вещей r, не соответствующее ожиданиям; • p отвергается, принимается q.
<p>(12) Ты тоже собираешься, <i>вместо того чтобы</i> готовить обеды, сидеть за компьютером? [Елена Павлова. Вместе мы эту пропасть одолеем! // «Даша», 2004]</p> <p>(13) Из голубого пластика и стекла, сверкая обтекаемыми изгибами, буфет напоминал по своим очертаниям <i>скорее</i> летательный аппарат, <i>чем</i> торговую точку. [Фазиль Искандер. Летним днем (1969)]</p>	
Исключение (14)	Исключение из рассмотрения ¹² (15)
<ul style="list-style-type: none"> • операция соотнесения элемента и множества; • пропозициональный уровень; • X содержит указание на множество P; • Y содержит указание на элемент q, исключаемый из этого множества. 	<ul style="list-style-type: none"> • операция соотнесения элемента и множества; • уровень высказывания; • X – информативный речевой акт, описывающий положение вещей p; • Y содержит указание на элемент q, который надо исключить, чтобы признать истинным p.
<p>(14) В первом классе любовь к учительнице ни к чему не привела, <i>кроме того, что</i> из-за нее я очень больно ударился головой о бревно. [С. Ткачева. День влюбленных... // «100% здоровья», 2003.01.15]</p> <p>(15) Едва ли кто-нибудь, <i>кроме</i> матери, заметил появление его на свет. [И. А. Гончаров. Обломов (1859)]</p>	
Экстенциональная генерализация (16)	Интенциональная генерализация ¹³ (17)
<ul style="list-style-type: none"> • операция соотнесения общего и частного; • пропозициональный уровень; • p – положение(я) вещей, имеющее(ие) место в некоторых обстоятельствах; • q – положение вещей, включающее p во множество $\{p_1, p_2... p_n\}$, а значит имеющее более широкий экстенционал. 	<ul style="list-style-type: none"> • операция соотнесения общего и частного; • метаязыковой уровень; • p – положение вещей, имеющее место в некоторых обстоятельствах; • q – обобщенное («без частных») представление положение вещей, сделанное на основании свойств p и имеющее более широкий экстенционал.

¹¹ Отсутствует в PDTB.

¹² Отсутствует в PDTB; подробнее см. [Инькова, Манзотти 2019].

¹³ Отсутствует в PDTB.

- (16) Никогда я не совал своего носа в литературу и в политику, не искал популярности в полемике с невеждами, не читал речей ни на обедах, ни на могилах своих товарищей... *Вообще* на моем ученом имени нет ни одного пятна. [А. П. Чехов. Скучная история (1889)]
- (17) Тогда шедший впереди откровенно вынул из-под пальто черный маузер, а другой, рядом с ним, отмычки. *Вообще*, шедшие в квартиру No 50 были снаряжены как следует. [М. А. Булгаков. Мастер и Маргарита (1929-1940)]

Экстенциональная спецификация ¹⁴ (18)	Интенциональная спецификация ¹⁵ (19)
<ul style="list-style-type: none"> • операции соотнесения общего и частного; • пропозициональный уровень; • X содержит обобщенное понятие или положение вещей <i>p</i>; • Y содержит более частное понятие <i>q</i>, сужающее экстенционал <i>p</i>. 	<ul style="list-style-type: none"> • операция соотнесения общего и частного; • метаязыковой уровень; • X содержит обобщенное понятие или положение вещей <i>p</i>; • Y содержит более частное понятие <i>q</i>, сужающее интенционал <i>p</i>.

- (18) Да и зачем оно, это дикое и грандиозное? Море, *например*? [И. А. Гончаров. Обломов (1848-1859)]
- (19) Подарил Марусе ценный, уникальный сувенир. *А именно* – конспиративную записку диссидента Шафаревича, написанную собственной рукой. [С. Довлатов. Иностранка (1986)]

7 Заключительные замечания

Определения, используемые в НБД, имеют, как мы видим, теоретическое значение, обладая объяснительной силой. Прежде всего, они позволяют увидеть семантическую близость некоторых ЛСО, а также тот факт, что разные ЛСО, но обладающие схожими свойствами, могут использоваться в схожих семантических структурах. Например, показатели метаязыковой альтернативы и перифрастического переформулирования могут взаимозаменяться, особенно в рамках простого предложения(20), без значительного изменения смысла, поскольку оба ЛСО установлены на метаязыковом уровне и уравнивают интенционалы *p* и *q*. А различия в семантике этих двух ЛСО позволяют их показателям сочетаться друг с другом не создавая тавтологии(21).

- (20) Можно утверждать, что сообщество есть письмо общества, *другими словами / или* – его «différence». [Е. В. Петровская. Безымянные сообщества (2010)]
- (21) Врач общей практики, *или, другими словами*, семейный врач, – это специалист с высшим медицинским образованием, имеющий юридическое право оказывать первичную многопрофильную медико-социальную помощь независимо от возраста и пола пациентов. [В. Шпикалов. В здравоохранении эксперименты недопустимы // «Восточно-Сибирская правда» (Иркутск), 2003.06.21]

Эта семантическая близость ЛСО проявляется и при сопоставительном анализе. Переводчики выбирают иногда показатель другого, но имеющего сходные семантические параметры ЛСО, чтобы передать семантику коннектора; см. в этом отношении перевод показателя экстенциональной генерализации *вообще* в (16) показателем ЛСО обобщающего переформулирования *bref* ‘корооче говоря’ в (22):

¹⁴ Подвид Exemplification ЛСО Instantiation в PDTB.

¹⁵ Подвид More detail ЛСО Level-of-Detail в PDTB.

- (22) Je n'ai jamais fourré le nez dans la littérature ni la politique, recherché la popularité en polémiquant contre des ignorants, prononcé de discours dans des dîners ou sur la tombe de mes confrères... *Bref*, mon nom de savant est sans tache [Tr. É. Parayre]

В определениях этих двух ЛСО есть общие элементы: оба они сигнализируют об обобщении, т.е. переходе к более широкому экстенсионалу.

В настоящее время ведется работа по интегрированию этих определений в НБД, что позволит автоматически исчислять семантическую близость ЛСО на основе наличия у них общих семантических признаков и их количества. Так, наличие общих признаков у ЛСО метаязыковой альтернативы и перифрастического переформулирования (+метаязыковой уровень, +тождество интенционала) позволяет говорить об их большей семантической близости, чем, например, для ЛСО перифрастического переформулирования и интенциональной спецификации, имеющих лишь один общий признак (+метаязыковой уровень).

Определения, используемые в НБД, позволяют, как нам кажется, определить на основе четких семантических критериев то или иное ЛСО, а значит преодолеть непоследовательность существующих классификаций и повысить уровень согласованности в работе экспертов и разметчиков, работающих в разных подходах и с разными языками, служа своего рода интерлингвой (о необходимости разработки такой интерлингвы см. [Sanders et al. 2018]). Это повышает и ценность аннотированного в разных подходах материала, который приобретает таким образом свойство reliability.

Литература

- [1] Alonso L., Castellon I., Padro L. (2002) Lexicón computacional de marcadores del discurso. *Procesamiento del Language Natural*. 2002. 29. P. 239–246.
- [2] Asher N. (1993) Reference to Abstract Objects in Discourse. Kluwer Academic Publishers: Dordrecht.
- [3] Asher N. (1996) L'interface pragmatique-sémantique et l'interprétation du discours [Pragmatic-semantic interface and discourse interpretation]. *Langages*. 1996. 123. P. 30–50.
- [4] Asher N., Lascarides A. (2003) Logics of Conversation. Cambridge: Cambridge University Press.
- [5] Benamara F., Taboada M. (2015) Mapping different rhetorical relation annotations: A proposal. *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (SEM 2015)*. Denver: ACL. P. 147–152.
- [6] Bras M. (2008) Entre relations temporelles et relation de discours. PhD Thesis. Toulouse University; URL: http://myriam.bras.free.fr/publis/HdR_Myriam.pdf.
- [7] Bunt H., Prasad R. (2016) ISO DR-Core (ISO 24617-8): Core Concepts for the Annotation of Discourse Relations. *Proceedings of the LREC 2016 Workshop "ISA-12: 12th Joint ACL – ISO Workshop on Interoperable Semantic Annotation"*. Bunt H. (ed.). Slovenia, Portorož, 2016. P. 45–54.
- [8] Carlson L., Marcu D. (2001) Discourse Tagging Reference Manual ISI Technical Report ISI-TR-545 54, 56; URL: <https://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>.
- [9] Chiarcos C. (2014) Towards interoperable discourse annotation. Discourse features in the Ontologies of Linguistic Annotation. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Calzolari N., Choukri Kh., Declerck Th., Loftsson H., Maegaard B., Mariani J., Moreno A., Odijk J., Piperidis S. (eds.). Reykjavik: ELRA. P. 4569–4577.
- [10] Demberg V., Scholman M. C. J., Asr F. T. (2019) How compatible are our discourse annotation frameworks? Insights from mapping RST-DT and PDTB annotations. *Dialogue & Discourse*. 2019. 1. P. 87–135.
- [11] Goncharov A. (2021) Classifications of intratextual relations: Bases and structuring principles [Klassifikatsiya vnutritekstovych otnosheniy: osnovaniya i principy strukturirovaniya]. *Voprosy Jazykoznanija*. 2021. 3. Pp. 97–119.
- [12] Grosz B. J., Sidner C. L. (1986) Attention, intentions, and the structure of discourse. *Computational Linguistics*. 1986. 3. P. 175–204.
- [13] Halliday M.A.K. (1985) An Introduction to Functional grammar. London: Edward Arnold.
- [14] Inkova O. (2019) Logical-semantic relations: classification problems [Logiko-semanticheskie otnosheniya: problemy klassifikatsii], O. Inkova, E. Manzotti, Text coherence: mereological logical-semantic relations [Svyaznost' teksta: mereologicheskie logiko-semanticheskie otnosheniya]. Moscow: Izdatel'skii Dom YaSK. Pp. 11–98.
- [15] Inkova O., Manzotti E. (2019) Text coherence: mereological logical-semantic relations [Svyaznost' teksta: mereologicheskie logiko-semanticheskie otnosheniya]. Moscow: Izdatel'skii Dom YaSK.
- [16] Knott A. (1996) A Data-Driven Methodology for Motivating a Set of Coherence Relations. PhD Thesis. University of Edinburgh.

- [17] Fraser B. (2009) An Account of Discourse Markers. *International Review of Pragmatics*. 2009. 1. P. 1–28.
- [18] Lang E. (1977) *Semantik der koordinativen Verknüpfung*. Berlin (DDR): Akademie. (The semantics of coordination. Amsterdam: John Benjamins, 1984.)
- [19] Mann W., Thompson S. (1988) Rhetorical structure theory: Towards a functional theory of text organization. *Text*. 1988. 8. P. 243–281.
- [20] Muller P., Vergez M., Prevot L., Asher N., Benamara F., Bras M., Le Draoulec A., Vieu L. (2012) Manuel d’annotation en relations de discours du projet ANNODIS [Annotation manual in discourse relations of the ANNODIS project]. *Carnets de grammaire, Rapport n°21*. CLLE-ERSS. Toulouse: Université de Toulouse Jean Jaurès.
- [21] Penn Discourse Treebank (PDTB) Project. University of Pennsylvania; URL: <https://www.seas.upenn.edu/~pdtb/>.
- [22] PDTB Research Group (2008). The Penn Discourse Treebank 2.0 Annotation Manual. Technical Report IRCS-08-01. Philadelphia: Institute for Research in Cognitive Science, University of Pennsylvania; URL: <https://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>.
- [23] PDTB Research Group (2019). The Penn Discourse Treebank 3.0 Annotation Manual, 2019; URL: <https://doi.org/10.35111/qebf-gk47>.
- [24] Polanyi L. (1995) *The Linguistic Structure of Discourse*. Stanford: CSLI Publications.
- [25] Prasad R., Webber B., Joshi A. (2017) The Penn Discourse Treebank: An Annotated Corpus of Discourse Relations. *Handbook of Linguistic Annotation*. Ide N., Pustejovsky J. (eds.). Dordrecht: Springer Science+Business Media. P. 1197–1217.
- [26] Reese B., Hunter J., Asher N., Denis P., Baldridge J. (2007) Reference Manual for the Analysis and Annotation of Rhetorical Structure (version 1.0). Technical Report. Austin: University of Texas, Departments of Linguistics and Philosophy; URL: <http://timeml.org/jamesp/annotation-manual.pdf>.
- [27] Roze C. (2013) *Vers une algèbre des relations de discours [Towards a Discourse Relation Algebra]*, PhD Thesis, Paris Diderot – Sorbonne Paris Cité University.
- [28] Sanders T., Demberg V., Hoek J., Scholman M., Asr F. T., Zufferey S., Evers-Vermeul J. (2018) Unifying Dimensions in Coherence Relations: How Various Annotation Frameworks are Related. *Corpus Linguistics and Linguistic Theory*; DOI: doi.org/10.1515/cllt-2016-0078.
- [29] Stede M., Umbach C. (1998) DiMLex: A lexicon of discourse markers for text generation and understanding. *Proceedings of the Joint 36th Meeting of the ACL and the 17th Meeting of COLING*. Montreal, August 10-14. P. 1238–1242.