

УДК 81'37; 81'42
DOI: 10.17223/19986645/74/5

О.Ю. Инькова

НЕСВЯЗНОСТЬ ТЕКСТА, ИЛИ О НЕКОТОРЫХ ПОДВОДНЫХ КАМНЯХ НА ПУТИ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТА

Рассматривается малоизученный вопрос анализа единиц, обеспечивающих членение информации на семантические блоки, интерпретируемые относительно заданного ими критерия (временного, пространственного и коммуникативного) и предлагает описание семантических свойств и текстовых функций рамочных выражений и их взаимодействия с другими сигналами, обеспечивающими адекватную интерпретацию текста. Этот аспект, как правило, не принимается во внимание в корпусах текстов, аннотированных с точки зрения параметров их связности.

Ключевые слова: связность текста, семантика, автоматическая обработка текста, риторические отношения, рамочные выражения

Вводные замечания

Большинство исследователей, занимающихся проблемами изучения текста, согласны с тем, что между составляющими его высказываниями существуют отношения связности, позволяющие превратить последовательность фраз в сообщение, имеющее смысл и коммуникативную цель. Несмотря на то, что проблемы связности текста занимают ученых разных областей знания уже более полувека, вопросы о том, что понимать под этими отношениями, какова их природа и т.д., продолжают вызывать дискуссии. Были сделаны многочисленные предложения, которые, однако, являются скорее инструментом семантического анализа текста, чем теоретической моделью его организации, поскольку основное внимание при анализе уделяется в первую очередь микроструктуре текста. Это справедливо, прежде всего, для представителей компьютерной лингвистики, которые несколько упрощенно – в силу принципа единственности, сформулированного У. Манном и С. Томпсон [1. Р. 248], – полагают, что между следующими друг за другом высказываниями существует только одно «риторическое отношение». Однако еще Дж. Хоббс в свое время подчеркивал, что при интерпретации текста читатель совершает параллельно несколько когнитивных операций, в частности интерпретирует смысл составляющих его слов, определяет коммуникативное намерение говорящего, устанавливает референтов анафорических выражений и то, в каких отношениях находятся между собой ситуации, описанные в тексте [2. Р.71]. Это значит, что связность текста устанавливается одновременно между взаимодействующими языковыми сущностями разной природы.

Принцип связанности, как он сформулирован У. Манном и С. Томпсон, постулирует также, что между всеми фрагментами текста должно быть обязательно некоторое «риторическое отношение» [1. Р. 248]. Создание корпусов текстов, аннотированных с точки зрения отношений связности, позволило, однако, заметить, что между соседними высказываниями может не существовать никакого «риторического отношения», что привело, в частности, разработчиков Penn Discourse Treebank (PDTB) к введению метки NoRel. Она проставляется в тех случаях, когда между смежными предложениями в пределах одного абзаца не может быть установлено никакого «дискурсивного отношения» [3. Р. 5]. Ср. (1):

(1) Jacobs Engineering Group Inc.'s Jacobs International unit was selected to design and build a microcomputer- systems manufacturing plant in County Kildare, Ireland, for Intel Corp. *Jacobs is an international engineering and construction concern. NoRel Total capital investment at the site could be as much as \$400 million, according to Intel* (пример из [4. Р. 25]).

В свою очередь, внимание к макроструктуре текста позволило увидеть, во-первых, что структуры более высокого уровня организации текста, например абзацы, оказывают влияние на интерпретацию структур более низкого уровня, а во-вторых, что в текстах присутствуют не только сигналы непрерывности сообщения, но и сигналы дискретности, предупреждающие о том, что между предыдущим и последующим контекстом нет непосредственной связи или что эта связь ослабляется (подробнее см. [5]). Эти сигналы играют не менее важную роль для правильной интерпретации текста, чем сигналы связности, устанавливая границы, в рамках которых должна быть интерпретирована информация, содержащаяся в микроединицах. Этот аспект семантического анализа текста остается пока малоизученным, поскольку доминирующим является восходящий подход к анализу текста, при котором основное внимание уделяется процессам интеграции минимальных дискурсивных единиц в единицы более высокого уровня; см., прежде всего, исследования, выполненные в рамках Теории риторической структуры (далее TRC).

Однако не менее продуктивным может быть нисходящий подход, ставящий во главу угла сегментирование текста на более мелкие единицы (ср., например, работы Ж.-М. Адама). Действительно, при интерпретации текста читатель одновременно осуществляет два процесса. Единицы, составляющие текст, расчленяются путем сегментации (типографически на письме, паузами, интонацией, движением глаз или головы в устной речи). Дискретность словесной цепочки начинается с разделения на слова, постоянного на письме и менее выраженного в устной речи; далее на графическом уровне сегментация продолжается делением на абзацы (в прозе) или строфы (поэтический текст) и делением на главы, части и т.д. в зависимости от длины текста. В то же время эти единицы текста, в частности благодаря соответствующим показателям, связываются друг с другом, создавая непрерывность текста на основе его дискретности. Поэтому в тексте важно различать два вида сигналов: «открывающие» показатели (ориентирован-

ные на то, что следует) и «закрывающие» показатели (ориентированные на то, что предшествует). Например, заглавная буква, начинающая предложение, и точка, его завершающая, открывающие и закрывающие скобки. К ним надо добавить полифункциональные показатели, например пробел или двойное тире (в отличие от тире, открывающего реплику прямой речи), которые могут быть как открывающими, так и закрывающими.

В структуре текста существует еще один уровень (рис. 1), промежуточный между макро- и микроструктурой, который, как правило, не принимается во внимание¹, но именно на этом уровне, который мы вслед за Ж.М. Адамом предлагаем называть мезоструктурным [6], происходят семантические процессы, играющие важную роль в понимании текста. Вертикальными стрелками обозначены процессы сегментирования в тексте, а горизонтальными – процессы интеграции микроединиц в более крупные единицы.

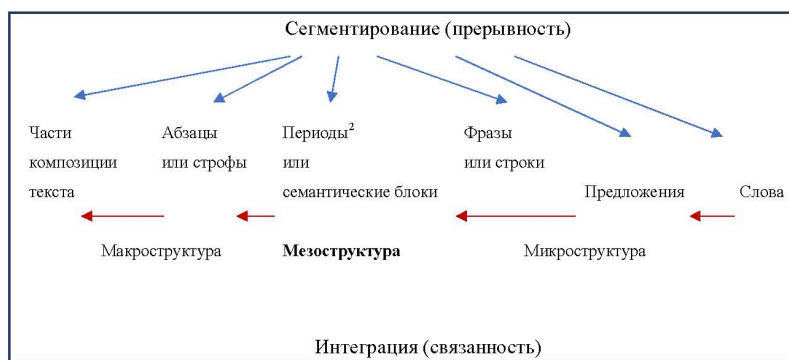


Рис. 1. Интеграция и сегментирование языковых единиц в тексте в подходе Ж.М. Адама

Взаимодействие мезо- и макроструктур, как и взаимодействие показателей связности, действующих на уровне этих структур, является довольно сложным, в частности потому, что границы единиц, относящихся к макро-, мезо- и микроструктуре, могут совпадать, а могут и пересекаться. Примером совпадения всех трех уровней можно считать мини-текст «Парикмахерская», где минимальная текстовая единица слово-предложение является одновременно и максимальной – текстом. С другой стороны, вряд ли можно, например, считать союз *а* в предложении 11 примера (2) «формальным маркером абзаца», единицы макроструктуры, в приводимом в [7. С. 148] примере, который мы даем в сокращении:

¹ Во многих аннотированных корпусах анализ дискурсивных отношений не выходит за пределы абзаца. Ср., в частности, приведенное выше определение NoRel в PDTB, в состав которого включен этот критерий.

² В русской грамматической традиции более распространен термин «сложное синтаксическое целое».

(2) 1. Аппарат набирал высоты, выше и выше, шел к снежным облакам, которые до горизонта обволокли небо плотным куполом. 2. Там, выше этих облаков, – яркое-яркое солнце, а внизу ослепительно белая пустыня – те же облака сверху. 3. Два мотора вертели два винта. 4. За их треском трудно было слушать друг друга пассажирам, которые сидели в каюте аппарата. 5. Они переписывались на клочках бумаги. <...> 8. «А мы все поднимаемся», – написал на бумажке веселый толстый пассажир, смотревший в окно, своему обалдевшему соседу. 9. Тот прочел, махнул раздраженно рукой, натянул еще глубже свою шляпу и ниже наклонился к полу. 10. Толстый пассажир достал из саквояжа бутерброды и принялся спокойно есть. 11. А впереди, у управления, сидели пилот, механик и ученик. 12. Все были тепло одеты, в кожаных шлемах. <...>

Считать союз *а* маркером абзаца значило бы, что каждый раз, когда мы употребляем этот союз, мы должны начинать новый абзац, что явным образом не соответствует языковой действительности. Об этом свидетельствует, в частности, предложение 2 в этом же тексте. В нем союз *а* употреблен в такой же семантической структуре сопоставления («там, выше этих облаков» – «а внизу») не просто внутри абзаца, а внутри сложного предложения.

И тем более можно считать, на наш взгляд, явным упрощением проблемы описания смысловой организации текста возможность применять тот же набор «дискурсивных отношений» к единицам его микро- и макро-структуры, что иногда ставится в заслугу ТРС.

Интеграция и сегментация единиц текста на мезоуровне

К мезоструктуре относятся явления семантического и макросинтаксического связывания. Само понятие мезоструктуры предполагает, что текст состоит не из последовательности предложений, а из упорядоченных фразовых объединений, семантических блоков, которые и определяют понимание текста [8. Р. 19]. Вопрос разработки теоретических принципов описания этих промежуточных единиц организации текста и их интерпретации, по мнению Ж.М. Адама, является центральным для сегодняшней лингвистики текста [6. Р. 22] и, добавим, его автоматической обработки и генерации. Как показывает рис. 1, мезоструктура находится на стыке микро- и макроструктуры и охватывает такие явления, которые уже не подчиняются строгим синтаксическим правилам и которые нельзя описывать в терминах валентностей, согласования и подобных синтаксических категорий, относящихся к микросинтаксису. Как известно, дискурс начинается именно там, где заканчивается действие этих правил, поэтому строгое определение элементов мезо- и макроструктуры текста вряд ли возможно¹. Т. ван Дейк так

¹ Заметим в этой связи, что даже определение предложения уже вызывает дискуссии, и определений, в зависимости от того, какой критерий кладется в его основу, дано более сотни, в том числе и в русской грамматической традиции. Ср., например, очень разные определения Ф.И. Буслаева, Д.Н. Овсяннико-Куликовского или Ф.Ф. Фортунатова.

предлагает определять упорядоченные фразовые блоки: «значение (или истина) данного предложения S зависит от значения (или истинности) упорядоченного кортежа предшествующих предложений» [8. Р. 20].

Показатели, действующие на мезоуровне, довольно разнородны, но их можно, не вдаваясь в детали, разделить на два класса. С одной стороны, языковые средства, которые сигнализируют о том, что между смежными или близкими сегментами текста можно установить некоторое отношение. Этот класс, в свою очередь, может быть разделен на два подкласса в зависимости от того, устанавливается ли это отношение между референтами (анафорические повторы) или между положениями вещей (коннекторы). С другой стороны, языковые единицы, указывающие на то, что несколько положений вещей должны быть интерпретированы одинаково относительно некоторого критерия, заданного этими единицами¹. Действительно, в любом тексте некоторые положения вещей, описывающие события, ситуации, мнения, оценки и т.д., имеют сферу действия, ограниченную, например, определенным временным отрезком (временной критерий) или местом (пространственный критерий). Автор текста вынужден также часто передавать информацию, которая считается обоснованной лишь некоторыми людьми или группой людей (коммуникативный критерий) и, как правило, сигнализирует об этих ограничениях, чтобы они были понятны читателю. Языковые единицы, задающие эти ограничения, мы предлагаем называть рамочными выражениями, или рамками, по аналогии с французским термином «cadre», английским «Framing Adverbials» или русским «рамка», использующимся в основном в сочетании «модальная рамка». Функция рамочных выражений заключается в том, что они задают критерий, с учетом которого должны быть интерпретированы попадающие в их сферу действия положения вещей.

Именно этим показателям, действующим на мезоуровне, и посвящено настоящее исследование. Оно не претендует на исчерпывающий характер, а лишь ставит проблему, пока еще мало исследованную при описании параметров, определяющих интерпретацию текста. В российской лингвистике работ, посвященных рамочным выражениям как в русском, так и в других языках, насколько нам известно, не существует. Наиболее разработан этот вопрос применительно к данным французского языка в исследованиях М. Шароля и его последователей; ср. основополагающую работу [12]. Редко принимается во внимание текстовая функция рамочных выражений и при автоматической обработке письменного текста. В силу своей синтаксически периферийной позиции они остаются за пределами синтаксической разметки (ср. деревья синтаксических зависимостей), а из-за того, что они не считаются показателями дискурсивных отношений, они остаются за пределами разметки дискурсивной (аннотирование с точки зрения «риторических» отношений).

¹ Ср. понятия «univers de discours» «дискурсивные миры» в [9] и «espaces mentaux» «ментальные пространства» в [10], а также теорию дискурсивных репрезентаций [11].

В связи с неразработанностью этой проблематики, в частности для русского языка, у лингвиста возникает целый ряд теоретических вопросов. Прежде всего, необходимо создать перечень языковых средств, способных образовывать рамки для объединения положений вещей в единый семантический блок, а также определить, каковы их виды в русском языке. Кроме того, следует выяснить функцию этих блоков в тексте и то, как взаимодействуют рамочные выражения и создаваемые ими семантические блоки, а также как взаимодействуют рамки с другими языковыми единицами, действующими на мезоуровне, в первую очередь, с коннекторами. Наконец, необходимо понять, как можно учитывать функцию рамочных выражений в аннотированных корпусах. На эти вопросы мы предлагаем ответы в следующих разделах.

3. Рамочные выражения: семантика, свойства, функции в тексте

Рассмотрим такой пример, в котором говорится о существующих в мире традициях встречать Новый год.

(3) **В Греции** существует обычай, согласно которому, ровно в полночь глава семейства выходит во двор и разбивает об стену плод граната. Если его зерна разлетятся по двору, в новом году семья будет жить счастливо. Отправляясь в гости, греки приносят с собой в подарок замшелый камень и оставляют его в комнате хозяев. Говорят: «Пусть деньги хозяев будут так же тяжелы, как этот камень». А **в Дании** лесники придумали отличный способ сохранить свои леса от браконьеров, желающих украсить свой дом лесной красавицей («Новогодние традиции в разных странах мира»: <http://welcomeworld.ru/spravochnik-puteshestvennika/traditsii-i-kharakter-narodov-mira/item/411-newyearstraditions/411-newyearstraditions>).

Обстоятельство места «в Греции», находящееся в начальной тематической позиции, задает пространственную рамку, т.е. определяет место описанных во всех следующих предложениях ситуаций, которые более никак не локализованы в этом абзаце. Эта пространственная рамка завершается введением новой рамки, также пространственной, «в Дании», которая определяет интерпретацию положений вещей, находящихся в ее правом контексте. Рамочные выражения «в Греции» и «в Дании» группируют таким образом последовательность предложений в единые блоки, указывая, что они должны быть интерпретированы относительно заданного им критерия и отдельно от соседних семантических блоков. При этом границы эти двух семантических блоков совпадают с разбиением текста на абзацы, а кроме того эти блоки соединены сопоставительным союзом *а*, который добавляется к семантическим отношениям, заданным рамочными выражениями. Заметим также, что единая временная рамка для всех положений вещей, описанных в тексте, занимающем несколько страниц, задается его названием «Новогодние традиции в разных странах мира». На этом основании и при отсутствии других указаний мы понимаем, например, обстоятельство «ровно в полночь» в первой фразе в (3) как «ровно в полночь новогодней ночи».

В (4) временную рамку задает обстоятельство «по воскресеньям», и благодаря ему универсальный квантификатор *всегда* понимается не как ‘всегда вообще’, а как ‘всегда по воскресеньям’.

(4) Отношения с Маяковским были ровными, он не был так ревнив и категоричен, и их жизнь на даче в Пушкине 1920–1921 годов ЛЮ вспоминала как самую спокойную и мирную. **По воскресеньям** приезжало много гостей. Аннушка *всегда* готовила котлеты. Если гостей было мало, давали две котлеты, если много – по одной (Василий Катанян. Лиля Брик. Жизнь (1999))¹.

Наконец, в (5) рамка «в принципе» задает обобщающий взгляд на вещи: сказанное может считаться верным, если не вдаваться в детали; следующая за ней рамка «с научной точки зрения» связывает истинность сказанного с определенной сферой деятельности: наукой.

(5) **В принципе**, идея «множества вселенных», идея «параллельных миров» находит все больше приверженцев среди астрономов. **С научной точки зрения** в ней нет ничего невозможного (Александр Грудинкин. Мир по «Планку» // Знание-сила. 2013).

В этой связи прежде всего возникает вопрос о том, какие именно языковые единицы могут создавать рамку для интерпретации событий. В работе [12] М. Шароль относит к таким единицам обстоятельства места и времени в начальной тематической позиции, синтаксически периферийной по отношению к предикату, а также вводные слова и выражения, соответствующие русским *с точки зрения X*, *согласно X*, *в принципе*, *вообще говоря*, показателям смены темы *что касается X* и др., не входящие в синтаксическую структуру предложения (рис. 2).

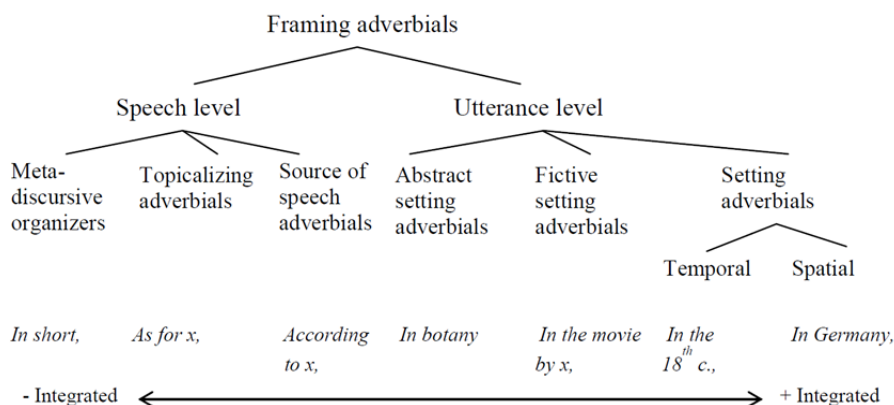


Рис. 2. Классификация рамочных выражений в подходе М. Шароля [14. Р. 6]

Общим для всех классов рамочных выражений является тематическая и синтаксически периферийная позиция, причем, как показано на рис. 2, сте-

¹ При отсутствии других указаний использованы примеры из Национального корпуса русского языка [13].

пень интеграции неодинакова у разных видов рамочных выражений. Наиболее интегрированными в синтаксическую структуру содержащего их высказывания являются обстоятельства места и времени, которые в русском языке, в отличие от французского, не отделяются запятыми. Наименее интегрированными – метатекстовые показатели типа *короче (говоря), другими словами*, которые, однако, в других подходах принято относить к коннекторам (см., например, [4–6]).

Термин «рамочные выражения» в таком понимании близок термину «детерминант», предложенному Н.Ю. Шведовой [15], но не получившему развития в российской лингвистике. Н.Ю. Шведова также обращает внимание на слабую синтаксическую зависимость детерминантов от предиката, говоря, вслед за В.В. Виноградовым, о связи «свободного, присоединительного характера» [15. С. 84]. Об этом, в частности, свидетельствует возможность заменить рамочные выражения времени и места на независимые синтаксические структуры. Ср. (3) выше и (6):

(6) **Колумбия.** Персонаж под названием Старый год провозглашается главным героем колумбийского карнавала. Он передвигается на ходулях и развлекает детей забавными историями (<https://evspb.ru/novyj-god-v-raznyx-stranax-samyie-interesnye-fakty/>).

Напротив, обстоятельства, находящиеся в рематической позиции или входящие в обязательную валентность предиката, свойством задавать рамку не обладают; ср. функционирование «по воскресеньям» в (7), в отличие от (4), и «в Париже» в (8), в отличие от (3):

(7) Он разведен, в прошлом браке остались дети. Теперь они приходят **по воскресеньям** в гости. Она добралась до чертогов Гименея далеко не девицей, к тому же и ребенок зачат вне брака (Евгения Пищикова. Пятиэтажная Россия (2007) // Русская Жизнь. 2008).

(8) Замечательный гербарий Александра Александровича Бунге (1803–1890) не был куплен Санкт-Петербургской академией и оказался **в Париже**. Многие типовые образцы российским ученым ныне приходится искать в Берлине, Женеве, Токио или Киото (обобщенный. Вести из институтов, лабораторий, экспедиций // Наука и жизнь. 2007).

Следующий вопрос, который требует решения, касается взаимодействия рамок двух смежных предложений. Возможны как минимум следующие варианты (подробнее см. [12]):

1) положение вещей *q* интегрируется в рамку¹, открытую положением вещей *p* и остающуюся открытой, как, например, в (3) или (4);

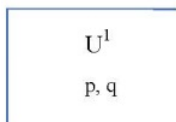


Рис. 3. Интеграция положений вещей в единой рамке

¹ На рисунке она обозначена заглавной буквой U («univers de discours» ‘дискурсивные миры’ в [9]).

2) положение вещей q интегрировано в рамку U^2 , закрывающую рамку U^1 , как в (5);

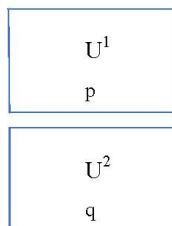


Рис. 4. Новая рамка

3) положение вещей q интегрировано в рамку U^2 , подчиненную рамке U^1 , как в (9) или (10).

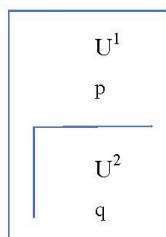


Рис. 5. Интеграция рамок

(9) **В Аргентине** по давней традиции служащие учреждений в последний день уходящего рабочего года выбрасывают из окон старые календари, ненужные ведомости и бланки. **В деловой части страны – Буэнос-Айресе** – уже к полудню тротуары и проезжая часть густо покрываются пухлым слоем бумаги. Никто не знает, *как и когда возник этот обычай. Не обходится без казусов.* Однажды не в меру разыгравшиеся сотрудники *одной из газет* выкинули за окно весь архив («Новогодние традиции в разных странах мира. URL: <http://welcomeworld.ru/spravochnik-puteshestvennika/traditsii-i-kharakter-narodov-mira/item/411-newyearstraditions/411-newyearstraditions>).

Рамочное выражение «в Аргентине» локализует положение вещей p = «по давней традиции служащие учреждений в последний день уходящего рабочего года выбрасывают из окон старые календари, ненужные ведомости и бланки». В следующем предложении устанавливается новая рамка «в деловой части страны – Буэнос-Айресе», она подчинена предыдущей, поскольку этот город, как известно, является столицей Аргентины. Следующие три предложения находятся в сфере действия этой рамки и соответствуют первой конфигурации. Таким образом, выбор между закрытием предыдущей рамки и подчинением ей зависит во многом от экстралингвистических знаний говорящих.

В (10) временная рамка «иногда» второго предложения встроена во временную рамку «по воскресеньям», открытую в первом предложении, т.е. ситуация «Лора читала вслух русскую газету» имеет место *иногда по воскресеньям*, а не просто *иногда*:

(10) **По воскресеньям** Алик и Лора долго завтракали, беседовали, курили. **Иногда** Лора читала вслух русскую газету. Проблемы, волновавшие эмигрантов, казались им надуманными (Сергей Довлатов. Третий поворот налево. 1984).

В связи с этим примером возникает также вопрос о том, как интерпретировать следующее предложение «Проблемы, волновавшие эмигрантов...»: как находящееся в сфере действия рамки «иногда (по воскресеньям)», в сфере действия более широкой рамки «по воскресеньям» или как находящееся вне этих рамок, т.е. проблемы казались надуманными вообще.

Не менее важным для анализа текста является вопрос взаимодействия рамок с дискурсивными отношениями, которые также устанавливаются на мезоуровне вторым классом показателей связности текста, коннекторами. Как мы видели, группировка предложений относительно некоторого критерия может накладываться на дискурсивные отношения, маркированные соответствующим показателем или нет. Здесь также возможны разные варианты соотношений сфер действия коннекторов и рамочных выражений. Положения вещей, связанные дискурсивным отношением, могут находиться в сфере действия единой рамки, как в (11):

(11) **На Руси** когда-то выпекали на Новый год из теста домашних животных: коней, коров, быков. *А когда* в дом приходили колядовать, гостей одаривали этими фигурками, разными сладостями, орехами. *Также* считали, что Новый год надо встречать в новом платье, обуви – *тогда* и весь год ходить в обновках. Обычно перед Новым годом отдавали все долги, прощали все обиды (Новогодние традиции в разных странах мира. URL: <http://welcomeworld.ru/spravochnik-puteshestvennika/traditsii-i-kharakter-narodov-mira/item/411-newyearstraditions/411-newyearstraditions>).

В пределах семантического блока, заданного обстоятельством места «на Руси», существует несколько логико-семантических отношений. В первом предложении – отношение спецификации, маркированное двоеточием: от обобщающего понятия «домашние животные» говорящий переходит к перечислению некоторых их видов. Второе предложение связано с первым отношением сопоставления, маркированным союзом *а*. В его правом контексте находится временное отношение, оформляемое союзом *когда*. Следующее высказывание присоединяется аддитивным отношением, маркируемым *также*, в правом контексте которого – отношение следствия, маркированное *тогда*, а в левом – оба высказывания, соединенные союзом *а*. Наконец, последнее высказывание связано с предыдущим также аддитивными отношениями, не имеющими показателя. В (3) выше ситуация несколько иная. Союз *а* устанавливает сопоставительные отношения между двумя положениями вещей, но как имеющими место в разных пространственных рамках.

В некоторых случаях, однако, именно рамочное выражение задает отношение, которым должны быть связаны положения вещей, и определяет границы фрагментов текста, связанных этим отношением. Так, в (12) рамочное выражение «с их точки зрения» определяет границы левого контекста последнего союза *но*: *p* = «она права в том, что ей не нравится строй жизни ее семьи и что ей хочется другого».

(12) Но из отвлеченных рассуждений и из других примеров она видела, что, **с их точки зрения**, [она права в том, что ей не нравится строй жизни ее семьи и что ей хочется другого.] *p* Но [теперь надо подчиниться и выждать, а потом устроиться по-своему.] *q* (Вера Микулич. Мимочка на водах. 1892).

При отсутствии рамки левым контекстом для *но* становится все предложение в целом, вводимое первым союзом *но*, как показывает трансформация (13). Снимается также и полифония говорящих.

(13) Но [из отвлеченных рассуждений и из других примеров она видела, что она права в том, что ей не нравится строй жизни ее семьи и что ей хочется другого.] *p* Но [теперь надо подчиниться и выждать, а потом устроиться по-своему.] *q*

М. Шароль [12. С. 18] приводит примеры, где рамочное выражение задает само дискурсивное отношение, которым должны быть связаны положения вещей. Ср. (14) с рамочным выражением и (15) без него, поскольку обстоятельство времени находится в позиции ремы и не задает рамки, замещенные у М. Шароля:

(14) **Le dimanche**, Paul va au café. Il ne veut pas y aller avec ses collègues de bureau. ‘**По воскресеньям** Поль ходит в кафе. Он не хочет туда ходить со своими коллегами по работе’ (перевод наш. – О.И.).

(15) Paul va au café **le dimanche**. Il ne veut pas y aller avec ses collègues de bureau. ‘Поль ходит в кафе **по воскресеньям**. Он не хочет туда ходить со своими коллегами по работе’ (перевод наш. – О.И.).

Вариант (14) с обстоятельством «le dimanche» ‘по воскресеньям’ в функции временной рамки может быть понят либо как (а) = ‘По воскресеньям Поль ходит в кафе, *но* он не хочет туда ходить (по воскресеньям) со своими коллегами по работе’, либо как (б) = ‘По воскресеньям Поль ходит в кафе, *потому что* он не хочет туда ходить (в остальные дни недели) со своими коллегами по работе’. В интерпретации (а) сфера действия рамочного выражения распространяется и на второе предложение, в (б) второе предложение находится вне сферы его действия. Вариант (15) может быть понят только как (в) = ‘Поль ходит в кафе по воскресеньям, потому что он не хочет туда ходить (вообще) со своими коллегами по работе’. Оба предложения толкуются *par défaut* как имеющие место вообще.

Таким образом, рамочные выражения играют важную роль в процессе интерпретации текста. Они не только группируют находящиеся в сфере их действия высказывания, которые должны быть интерпретированы относительно некоторого единого критерия, но и могут определять, какие дискурсивные отношения следует установить между высказыванием, находящимся в их сфере действия, и последующими.

4. Рамочные выражения в аннотированных корпусах

В различных теориях дискурсивных отношений и созданных на их основе аннотированных корпусах текстов группировка положений вещей в семантические блоки, задаваемые рамочными выражениями, как правило, не учитывается. Причин на то несколько. Прежде всего, в некоторых корпусах, например в Penn Discourse Treebank (www.seas.upenn.edu/~pdtb) или в Надкорпусной базе данных коннекторов (<http://a179.frccsc.ru/RSCF10004/main.aspx>), предметом аннотирования являются логико-семантические отношения, т.е. только те, которые потенциально могут быть выражены вторым классом единиц, обеспечивающим связность текста на мезоуровне, а именно коннекторами. В корпусах, опирающихся на подходы, в которых дискурсивные отношения понимаются более широко, в частности в тех, которые опираются при аннотировании на ТРС (см., например, RST Discourse Treebank, [16]), тип связности, обеспечивающийся рамочными выражениями, тоже не учитывается. Это связано, с одной стороны, с тем, что минимальной дискурсивной единицей в подавляющем большинстве подходов считается клауза, тогда как рамочные выражения в своем большинстве являются непредикативными структурами, а некоторые из них входят в пропозициональное содержание предложения, в составе которого они находятся. С другой стороны, связность текста, обеспечиваемая рамочными выражениями, и связность, в основе которой лежат отношения, выражаемые коннекторами, часто накладываются друг на друга; см. (3) выше. Ситуация усложняется и тем, что некоторые рамочные выражения, указывающие, например, на источник информации, выполняют в тексте ту же функцию, что и глаголы речи (16), и размечаются в аннотированных согласно ТРС корпусах как вводящие сателлит отношения Attribution (17). Однако этот случай специально оговаривается, причем, как правило, только для *according to*¹:

(16) По данным аналитического центра «Индикаторы рынка недвижимости IRN.RU», средняя стоимость квадратного метра жилья в Фили-Давыдовке составляет 171 тыс. руб., в Кунцеве – 173 тыс., в Тропареве-Никулине – 178 тыс., в районе Филевский парк – 187 тыс., в Крылатском – 203 тыс., в районе Проспект Вернадского – 210 тыс. руб. Средняя цена однокомнатной квартиры в панельном доме в указанных районах – около 7,3 млн руб., двухкомнатной – 9,5 млн руб., сообщили «Газете.Ру» в компании «Пересвет-Инвест» (URL: <https://www.irn.ru/articles/36629.html>).

(17) [September orders for machine tools rebounded from the summer doldrums, but remained 7.7% below year-earlier levels,]N [according to figures from NMTBA – the Association for Manufacturing Technology.]S – Attribution (wsj_628: 4-5/6-7) (пример из [17. P. 25]).

¹ “When the source of the attribution is a phrase, beginning with the expression *according to*, rather than a clause, it should also be marked as an elementary discourse unit” [16. P. 8].

Мы видели, однако, что в случаях типа (12), когда ядро отношения Attribution содержит рамочное выражение, источник содержащейся в нем информации может не совпадать с источником, указанным в сателлите: в примере «Она видела, что она права» $\neq$ «Она видела, что она права с их точки зрения».

Что касается пространственных и временных рамок, то они не принимаются во внимание при разметке согласно TPC, поскольку они не имеют пропозициональной структуры и не выделяются, следовательно, в самостоятельные дискурсивные единицы. Исключение составляют случаи типа (6), т.е. когда временная или локализирующая рамка оформлены в независимое предложение. В корпусе RST Discourse Treebank [16, 17] такие семантические конфигурации относятся не к собственно дискурсивным отношениям, а к организации текста. Для них предлагается следующая разметка:

(18) [Friday, October 27, 1989]^{DATE} [The key U.S. and foreign annual interest rates below are a guide to general levels but don't always represent actual transactions...]^{TEXT} wsj_0602 (пример из [17. P. 72]).

(19) [CHICAGO:]^{POINT-OF-ORIGIN} [Sears, Roebuck & Co. is struggling as it enters the critical Christmas season...]^{TEXT} wsj_1105 (пример из [17. P. 73]).

Иной подход представлен в аннотированном корпусе ANNODIS, разработанном в рамках SDRT Н. Ашера [18]. В первоначальной версии SDRT вопрос о сегментировании текста на минимальные дискурсивные единицы не обсуждался, а за основу по умолчанию был принят все тот же пропозициональный подход (минимальная единица = клауза). Однако в последующих работах [19, 20] было показано, что рамочные выражения обладают определенной дискурсивной самостоятельностью: их сфера действия может выходить за рамки предложения, в котором они находятся. Это привело к пересмотру критерия выделения минимальной дискурсивной единицы: в его основу легло наличие у сегмента текста, претендующего на такой статус, дискурсивной функции независимо от его синтаксической природы [21]. Это позволяет унифицировать разметку непропозициональных показателей отношения Attribution и других видов рамочных выражений, для которых в корпусе ANNODIS предлагается проставлять дискурсивное отношение Encadrement 'Рамка' (при разметке используется английский термин Frame; рис. 6).

[En médecine,]_10 [la psychothérapie est rattachée à la psychiatrie,]_11 [et [en psychologie,]_12 elle relève d'une formation post-graduée.]_13
 continuation(10,12), frame(10,11), frame(12,13)

Рис. 6. Пример разметки рамочных выражений в корпусе ANNODIS [22. P. 21]

В приведенном на рис. 6 высказывании использованы два рамочных выражения, задающие области знаний, для которых верно сказанное в попадающих в их сферу действия предложениях: «en médecine» 'в медицине' (сегмент 10) и «en psychologie» 'в психологии' (сегмент 12). Отношение

Continuation ‘Продолжение’, вводимое союзом *et* ‘и’, устанавливается между положениями вещей «психотерапия является частью психиатрии» (фрагмент 11) и «она преподается как часть специализации» (фрагмент 13), находящимися в сфере действия соответствующих рамок: фрагмент 10 («*en médecine*») и фрагмент 12 («*en psychologie*»), о чем говорят номера сегментов текста, связанных этим отношением: Continuation (10, 12).

Поскольку примеров аннотирования в работе [23] приводится довольно мало и не на все случаи, остается неясным, как должны быть размечены высказывания типа (12), когда фрагмент текста содержит одновременно отношение Attribution и рамочное выражение, которые вступают в своего рода конфликт: содержание главного предложения приписывается глаголом *видеть* «ей» (Bave), а на самом деле эта точка зрения принадлежит другим субъектам, к которым отсылает местоимение *их*. По-видимому, разметка должна иметь следующий вид:

(20) Но [из отвлеченных рассуждений и из других примеров она видела,]_1 что, [с **их точки зрения**,]_2 [она права в том, что ей не нравится строй жизни ее семьи]_3 [и что ей хочется другого.]_4 [*Но* теперь надо подчиниться и выждать,]_5 [а потом устроиться по-своему.]_6 Attribution(1, 2), Frame(2, 3–6), Continuation(3, 4), Temp(5, 6), Contrast(3–4, 5–6).

Согласно такой разметке отношение Attribution устанавливается непосредственно с рамочным выражением, а не с пропозицией, входящей в сферу его действия, что позволяет отразить полифонию говорящих и представить более адекватно семантику данного высказывания.

С другой стороны, нельзя не подчеркнуть, что рамочные выражения и коннекторы имеют разную семантическую природу и особенности функционирования и часто вступают в довольно сложное взаимодействие, поэтому разметка дискурсивных функций языковых средств этих двух классов может потенциально привести к конфликту при аннотировании.

Заключительные замечания

Анализ средств связывания на мезоуровне, промежуточном между микро- и макроструктурой текста, показал, насколько важно учитывать их взаимодействие при интерпретации текста, а значит, и при его аннотировании. Это еще раз подчеркивает справедливость гипотезы Дж. Хоббса о том, что понимание текста складывается в результате одновременной интерпретации языковых единиц разных уровней и разной семантической природы, и ставит под вопрос правомерность утверждения, что между единицами всех уровней иерархии текста существуют дискурсивные отношения одного и того же типа. Если говорить о двух классах единиц мезоуровня – коннекторах и рамочных выражениях, то в отличие от коннекторов, связывающих фрагменты текста, рамочные выражения группируют находящиеся в их правом контексте высказывания и в то же время сигнализируют, какие единицы должны быть интерпретированы относительно заданного им единого критерия, а какие – нет. Смена рамки указывает на

то, что между предыдущим семантическим блоком и блоком, открываемым новой рамкой, нет непосредственной связи, что новый семантический блок должен быть интерпретирован отдельно от предыдущего и относительно другого критерия. Представляет, таким образом, интерес не только изучение процессов интеграции минимальных дискурсивных единиц в большие (т.е. восходящий подход), но и процесс членения макроединиц на меньшие (т.е. нисходящий подход). В восходящем подходе рамки будут считаться сигналом интеграции высказываний в более крупную дискурсивную единицу, в нисходящем подходе – сигналом локального прерывания связности, началом новой дискурсивной единицы. Применение нисходящего подхода ставит, однако, перед лингвистами целый ряд вопросов, на которые еще предстоит найти ответы: как мы интерпретируем ситуации при отсутствии рамок, как взаимодействуют их сферы действия при наличии в тексте нескольких рамок, каковы языковые средства, сигнализирующие завершение сферы действия рамочного выражения (в частности, какую роль в закрытии рамки играют абзацы и другие единицы макроуровня).

Наконец, в том, что касается аннотирования текстов, следует, по-видимому, признать, учитывая разнородность языковых единиц, участвующих в организации текста и обеспечивающих его связность, что каждый разряд этих единиц должен размечаться отдельно и лишь затем могут быть показаны механизмы их взаимодействия.

Литература

1. Mann W.C., Thompson S.A. Rhetorical Structure Theory: Towards a Functional Theory of Text Organization // *Text & Talk*. 1988. № 8 (3). P. 243–281.
2. Hobbs J.R. Coherence and Coreference // *Cognitive science*. 1979. № 3. P. 67–90.
3. Webber B., Prasad R., Lee A., Joshi A. The Penn Discourse Treebank 3.0. Annotation Manual. 2019. URL: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.
4. PDTB Research Group. The Penn Discourse Treebank 2.0 Annotation Manual. Technical Report IRCS-08-01. Philadelphia : Institute for Research in Cognitive Science, University of Pennsylvania, 2008. URL: <https://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>.
5. Инькова О. Логико-семантические отношения: проблемы классификации // Инькова О., Манзотти Э. Связность текста: мереологические логико-семантические отношения. М., 2019. С. 11–98.
6. Adam J.-M. éd. Faire texte. Frontières textuelles et opérations de textualisation. Besançon : Presses universitaires de Franche-Comté, 2015. 356 p.
7. Кибрик А.А. Дискурс // Кибрик А.Е. и др. Введение в науку о языке. М., 2019. С. 126–163.
8. Dijk T.A. van. Text grammar and text logic // *Studies in Text Grammar* / eds by J. Petöfi, H. Rieser. Dordrecht : Reidel, 1973. P. 17–76.
9. Martin R. La logique du sens. Paris : PUF, 1983.
10. Fauconnier G. Espaces mentaux. Paris : Minuit, 1984.
11. Kamp H., Reyle U. From Discourse to Logic. Dordrecht : Kluwer, 1993.
12. Charolles M. L'encadrement du discours // *Cahier de Recherche Linguistique*. 1997. № 6. P. 1–73. URL: <https://hal.archivesouvertes.fr/hal-00665849>

13. Национальный корпус русского языка. URL: www.ruscorpora.ru
14. Sarda L., Carter-Thomas Sh., Fagard B., Charolles M. Adverbials in Use. From Predicative to Discourse Functions. Louvain-La-Neuve : Mons, Presses Universitaires de Louvain, 2014.
15. Шведова Н.Ю. Детерминирующий объект и детерминирующее обстоятельство как самостоятельные распространители предложения // Вопросы языкознания. 1964. № 6. С. 77–93.
16. Carlson L., Marcu D., Okurovsky M.E. RST Discourse Treebank. Philadelphia, Linguistic Data Consortium. 2002. URL: <https://catalog.ldc.upenn.edu/LDC2002T07>
17. Carlson L., Marcu D., Okurovsky M.E. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory // Proceedings of the Second SIGdial Workshop on Discourse and Dialogue. 2001. URL: <https://www.aclweb.org/anthology/W01-1605>
18. Asher N., Lascarides A. Logics of conversation. Cambridge : Cambridge University Press, 2003.
19. Charolles M., Le Draoulec A., Péry-Woodley, Sarda L. Temporal and spatial dimensions of discourse organization // Journal of French Language Studies. 2005. Vol. 15. P. 203–218.
20. Vieu L., Bras M., Asher N., Aurnague M. Locating adverbials in discourse // Journal of French Language Studies. 2005. Vol. 15. P. 173–193.
21. Prévot L., Vieu L., Asher N. Une formalisation plus précise pour une annotation moins confuse : la relation d'élaboration d'entité // Journal of French Language Studies. 2009. Vol. 19. P. 207–228.
22. Muller P., Vergez M., Prevot L., Asher N., Benamara F., Bras M., Le Draoulec A., Vieu L. Manuel d'annotation en relations de discours du projet ANNODIS. (Carnets de grammaire, Rapport n°21. CLLE-ERSS.) Toulouse : Université de Toulouse Jean Jaurès, 2012.

Text Incoherence, or Some Pitfalls of Automatic Text Processing

Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology. 2021. 74. 81–98. DOI: 10.17223/19986645/74/5

Olga Yu. Inkova, University of Geneva (Geneva, Switzerland), Institute of Informatics Problems of the Russian Academy of Sciences (Moscow, Russian Federation). E-mail: olga.inkova@unige.ch

Keywords: text coherence, semantics, automatic text processing, rhetorical relations, frame expressions.

The article discusses the little-studied aspects of analysis of framework expressions, i.e. units that divide information into semantic blocks interpreted according to certain criteria (temporal, spatial, or communicative) set by framework expressions thus ensuring text coherence. The analysis is based on examples from the Russian National Corpus. Unlike most studies using the ascending approach to text coherence, i.e. the integration of minimal discourse units higher-level units, such as in the Rhetorical Structure Theory, this article uses a descending approach, analyzing, on the contrary, the segmentation of the text into smaller units. This approach proves its productivity by allowing to show that texts have not only signals of coherence, but also signals of discreteness, warning that there is no direct connection between the previous and the subsequent context or that there is no such connection at all. Frame expressions function as those signals. The article, without claiming to be exhaustive, raises questions that arise while describing framework expressions; the author gives answers to some of them. First of all, semantic and functional properties of framework expressions are described: 1) weak syntactic dependence on the predication (peripheral syntactic position), 2) topic status, 3) certain semantics features. The author then analyzes how scopes of frame expressions interact and gives a number of possible configurations: 1) the state of affairs q is integrated into the U_1 frame that opens with the state of affairs p and

remains open; 2) the state of affairs q is integrated into the U_2 frame that closes the U_1 frame; 3) the state of affairs q is integrated into the U_2 , a frame that is subordinate to the U_1 frame. Finally, the author analyzes the interaction of frame expressions with connectives as logical-semantic relations markers. This interaction is manifested in the fact that the frame defines the boundaries of text spans between which this relation is established. The final section shows, using four resources as examples (Penn Discourse Treebank, Supra-corpora database of connectives, RST Discourse Treebank, ANNODIS) the way in which frame expressions are used in the text annotation process. Establishing the heterogeneity of the linguistic units ensuring text coherence, the author concludes that each category of these units should be annotated separately and only in that case the mechanisms of their interaction can be shown. The results obtained can be used to study the discourse structure and in text annotation.

References

1. Mann, W.C. & Thompson, S.A. (1988) Rhetorical Structure Theory: Towards a Functional Theory of Text Organization. *Text & Talk*. 8 (3). pp. 243–281.
2. Hobbs, J.R. (1979) Coherence and Coreference. *Cognitive Science*. 3. pp. 67–90.
3. Webber, B., Prasad, R., Lee, A. & Joshi, A. (2019) *The Penn Discourse Treebank 3.0. Annotation Manual*. [Online] Available from: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.
4. 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. [Online] Available from: <https://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>.
5. In'kova, O. (2019) Logiko-semanticheskie otnosheniya: problemy klassifikatsii [Logical-semantic relations: problems of classification]. In: In'kova, O. & Manzotti, E. *Svyaznost' teksta: mereologicheskie logiko-semanticheskie otnosheniya* [Text connectivity: mereological logical-semantic relations]. Moscow: YaSK. pp. 11–98.
6. Adam, J.-M. (ed.) (2015) *Faire texte. Frontières universitaires de Franche-Comté. textualisation*. Besançon: Presses universitaires de Franche-Comté.
7. Kibrik, A.A. (2019) Diskurs [Discourse]. In: Kibrik, A.A. et al. *Vvedenie v nauku o yazyke* [Introduction to the science of language]. Moscow: Buki-Vedi. pp. 126–163.
8. van Dijk, T.A. (1973) Text grammar and text logic. In: Petöfi, J. & Rieser, H. (eds) *Studies in Text Grammar*. Dordrecht: Reidel. pp. 17–76.
9. Martin, R. (1983) *La logique du sens*. Paris: PUF.
10. Fauconnier, G. (1984) *Espaces mentaux*. Paris: Minuit.
11. Kamp, H. & Reyle, U. (1993) *From Discourse to Logic*. Dordrecht: Kluwer.
12. Charolles, M. (1997) L'encadrement du discours. *Cahier de Recherche Linguistique*. 6. pp. 1–73. [Online] Available from: <https://hal.archivesouvertes.fr/hal-00665849>
13. *Russian National Corpus*. [Online] Available from: www.ruscorpora.ru (In Russian).
14. Sarda, L. et al. (2014) *Adverbials in Use. From Predicative to Discourse Functions*. Louvain-La-Neuve: Mons, Presses Universitaires de Louvain.
15. Shvedova, N.Yu. (1964) Determiniruyushchiy ob'ekt i determiniruyushchee ob'stoyatel'stvo kak samostoyatel'nye rasprostraniteli predlozheniya [Determining Object and Determining Circumstance as Independent Distributors of the Sentence]. *Voprosy yazykoznaniya*. 6. pp. 77–93.
16. Carlson, L., Marcu, D. & Okurovsky, M.E. (2002) *RST Discourse Treebank. Philadelphia, Linguistic Data Consortium*. [Online] Available from: <https://catalog.ldc.upenn.edu/LDC2002T07>
17. Carlson, L., Marcu, D. & Okurovsky, M.E. (2001) Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*. [Online] Available from: <https://www.aclweb.org/anthology/W01-1605>

18. Asher, N. & Lascarides, A. (2003) *Logics of conversation*. Cambridge: Cambridge University Press.
19. Charolles, M. et al. (2005) Temporal and spatial dimensions of discourse organization. *Journal of French Language Studies*. 15. pp. 203–218.
20. Vieu, L. et al. (2005) Locating adverbials in discourse. *Journal of French Language Studies*. 15. pp. 173–193.
21. Prévot, L., Vieu, L. & Asher, N. (2009) Une formalisation plus précise pour une annotation moins confuse: la relation d'élaboration d'entité. *Journal of French Language Studies*. 19. pp. 207–228.
22. Muller, P. et al. (2012) *Manuel d'annotation en relations de discours du projet ANNODIS*. (Carnets de grammaire, Rapport n°21. CLLE-ERSS.) Toulouse: Université de Toulouse Jean Jaurès.