

Е. А. Сулейманова

Семантический анализ контекстных дат

Аннотация. В статье предлагается формальный аппарат для представления семантического значения контекстных дат, основанный на лингвистической модели календарного времени. Контекстные даты — случаи отсылочной и эллиптической референции к единичным календарным интервалам в тексте — входят в состав значительной части т. наз. недоопределенных указаний на время. В терминах лингвистической модели представляется семантическое значение структурных элементов контекстных дат, содержащих индексальные (отсылочные) показатели, с учетом типа индексального показателя, типа элемента и его позиции в контекстной дате. Рассмотрены различные виды эллипсиса в контекстных датах и способы восстановления опущенных значений.

Ключевые слова и фразы: автоматическая обработка текста, интерпретация недоопределенных указаний на время, семантический анализ.

Введение

Неотъемлемая составляющая задачи извлечения фактов — локализация во времени события, соответствующего извлекаемому факту. Информация, позволяющая локализовать событие на временной оси, передается посредством разнообразных по форме и содержанию текстовых выражений — временных указателей, или темпоральных выражений.

В качестве ключевого элемента в большинство временных указателей входят выражения, обозначающие (имеющие своим референтом) конкретный интервал на календарной оси. Именно они позволяют датировать события с той или иной степенью точности. Для таких выражений нами был введен общий термин *текстовая дата* [1].

Работа выполнена в рамках НИР «Моделирование модально-временного аспекта описания ситуаций в задаче извлечения информации из текстов», номер гос. регистрации 01201455353.

© Е. А. Сулейманова, 2015

© Институт программных систем имени А. К. Айламазяна РАН, 2015

© Программные системы: теория и приложения, 2015

Наряду с полными, *абсолютными*, текстовыми датами (содержащими в явном виде все необходимые для идентификации календарного интервала значения, например *01.04.2001*, *1 апреля 2000 года*, *весной 2001 года*), в новостных текстах часто встречаются случаи не прямой (отсылочной, индексальной) и неполной (эллиптической) референции к календарным интервалам: *позапрошлый год*, *этот квартал*, *тот же месяц прошлого года*, *понедельник*, *сегодня* и т. п.¹ Определить интервал, к которому отсылает такое выражение, невозможно без обращения к той или иной форме контекста. Такие случаи мы объединяем под общим названием *контекстные даты*.

Конечной целью анализа текстовой даты при извлечении темпоральной информации является идентификация соответствующего ей временного интервала в формальной модели времени (онтологии времени). Если речь идет об абсолютной текстовой дате, то эта задача решается относительно просто: достаточно распознать значения-идентификаторы в составе даты и нормализовать их в соответствии со спецификациями формального языка (языка онтологии времени). Всё разнообразие абсолютных текстовых дат сводится к множеству типовых — цифровых, текстовых или смешанных — форм записи, которые могут быть обработаны правилами на основе сопоставления с шаблонами (паттернами).

Если провести параллель с именной референцией, то абсолютные текстовые даты — это полные имена собственные календарных интервалов (*5 апреля 1999 года*, *второй квартал 1999 года*), тогда как контекстные даты — это либо усеченные варианты собственных имен, требующие — для идентификации интервала — восстановления до полного имени (*5-го*, *второй квартал*), либо определенные дескрипции (*прошлая пятница*, *тот же месяц прошлого года*, *предшествующий квартал*), в т.ч. дейктические или анафорические местоименные наречия (*сегодня*, *накануне*). В дискурсе каждая такая дескрипция имеет своим референтом конкретный интервал. Но кроме этого, каждая такая дескрипция имеет собственное, инвариантное относительно дискурса, семантическое значение, например:

накануне = '(в) день, предшествующий дню X', где X — день, упомянутый в предшествующем фрагменте, или день, содержащий время упомянутого ранее события;

¹Приводятся данные о том, что относительные и неполные указания на время составляют в среднем 54% от всех размеченных темпоральных выражений в новостных текстах, 32% — в текстах на исторические темы и 25% — в текстах клинических записей [4].

тот же квартал прошлого года = 'квартал, имеющий такой же идентификатор, что и квартал X , но содержащийся в году, предшествующем году Y ', где X — идентификатор квартала, упомянутого в предшествующем фрагменте или содержащего интервал, упомянутый в предшествующем фрагменте; Y — год, содержащий момент речи.

Таким образом, если абсолютная текстовая дата связана со своим денотатом (календарным интервалом) напрямую, то у контекстной даты эта связь опосредована семантическим значением, включающим в себя индексальную составляющую (отсылки к текстуальному или прагматическому контексту).

Задачу идентификации (вычисления) интервала контекстной даты можно представить в виде трех шагов:

- (1) Семантический анализ контекстной даты, результатом которого является формализованное представление ее значения, которое (представление) может быть передано в формальную модель времени и проинтерпретировано как инструкция для выполнения.
- (2) Разрешение анафорических и дейктических отсылок в составе контекстной даты с целью обнаружения, в текстуальном или прагматическом контексте, данных для подстановки в инструкцию.
- (3) Вычисление значений контекстной даты в формальной модели времени.

Две первые задачи — это задачи лингвистического уровня. Их решение должно опираться на некоторую концептуальную модель, которая, с одной стороны, ориентирована на языковые способы выражения временной информации, а с другой стороны — гарантирует, что построенное в ее терминах семантическое представление будет однозначно интерпретируемо в формальной модели.

Современные западные подходы к проблемам автоматической обработки темпоральной информации имеют в основном прагматическую направленность и ориентированы прежде всего на максимально широкий охват разнообразных текстовых явлений. В контексте задачи, формулируемой как распознавание и нормализация темпоральных выражений (Temporal Expression Recognition and Normalization), под интерпретацией темпорального выражения понимается приведение его в соответствие общепризнанному стандарту представления временной информации TimeML [2] — т.е. представление на языке формальной онтологии времени. Понятие контекстной даты в задачах извлечения темпоральной информации не выделяется, а само явление относится

к сфере т. наз. недоопределенных темпоральных выражений (тип *Underspecified Temporal Expressions* в разметке TИМEX3 стандарта TimeML; к нему, наряду с *Monday, last year*, относят также и конструкции типа *two years ago*), которые противопоставляются абсолютным, или полностью определенным (*Fully Specified Temporal Expressions*). Иногда различают относительные (*relative*) и неполные (*incomplete*), или частично определенные (*partially specified*), темпоральные выражения [3, 4]. Задача нормализации недоопределенных указаний на время признается очень сложной. При этом в описаниях конкретных решений вопросы формального представления семантического значения темпоральных выражений чаще всего остаются без внимания: переход от поверхностного выражения к формально-онтологическому представлению обеспечивается встроенными процедурами и представлениями *ad hoc*, которые иллюстрируются на единичных примерах. С другой стороны, формально-семантические подходы к интерпретации временной информации [5], имеющие под собой прочный теоретический фундамент, плохо масштабируются для практических задач. В связи с этим некоторые разработчики прибегают к компромиссным стратегиям, выстраивая в своих практически ориентированных решениях некоторый промежуточный формальный уровень представления контекстно-независимой семантики недоопределенных выражений (локальная семантика в терминах рекурсивных матриц атрибут-значение [6], представление значений временных выражений на языке лямбда-исчисления [7] или с помощью системы временных (*temporaxy*) атрибутов [8]).

Немногочисленные работы, посвященные автоматическому анализу темпоральной информации в текстах на русском языке [9–11], не ставят своей целью интерпретацию временных выражений в указанном выше смысле.

В настоящей статье предлагается формальный аппарат для представления семантического значения текстовых указаний на время, основанный на лингвистической модели календарного времени. В терминах лингвистической модели описывается семантическое значение структурных элементов контекстных дат, содержащих индексальные (отсылочные) показатели, с учетом типа индексального показателя, типа элемента и его позиции в текстовой дате. Рассмотрены различные виды эллипсиса в контекстных датах и способы восстановления опущенных значений.

Лингвистическая модель времени не предназначена для формальных вычислений. Ее цель — обеспечить передачу в формальную модель

(онтологию времени) в унифицированном виде всей доступной из текста информации, необходимой для таких вычислений.

1. Лингвистическая модель календарного времени

Предлагаемая лингвистическая модель может быть названа *наивной*² моделью календарного времени. Это модель интервального типа (понятия «точка» в модели нет). Модель не включает представление времени суток.

1.1. Календарные интервалы

Основная календарная шкала S — это временная шкала с единицей, представляющей собой интервал одного из следующих типов: *день/сутки, декада, месяц, квартал, полугодие, год, десятилетие, век, тысячелетие, эра*.

Наряду с основной шкалой будем использовать:

- шкалу недель с единицей — интервалом типа *неделя*;
- шкалу дней недели с единицей — интервалом типа *день недели*;
- шкалу с единицей — интервалом типа *время года* (имеются в виду четыре календарных времени года продолжительностью по три месяца).

Единичный календарный интервал типа S — единичный интервал на любой из календарных кал.

Единица длительности S — длительность единичного календарного интервала типа S .

Для идентификации единичных календарных интервалов используются числовые и лексические идентификаторы (названия). Ни один из типов календарных интервалов не обладает идентификаторами, которые однозначно определяли бы интервал на всей оси времени (за исключением условного типа *эра*, имеющего два уникальных значения).

Интервал типа *день/сутки* имеет числовой идентификатор (число месяца — от 1 до 31), уникальный только в пределах периода — интервала типа *месяц*.

²Если понимать «наивный» в том смысле, в котором его использует Ю. Д. Апресян, называя «наивной моделью мира» «образ мира, запечатленный в языке», [...] «во многих существенных деталях» отличающийся «от научной картины мира» [12].

Интервал типа *декада* имеет числовой идентификатор (от 1 до 3), уникальный в пределах периода — интервала типа *месяц*.

Интервал типа *месяц* имеет идентификаторы обоих типов — название месяца и числовой идентификатор (от 1 до 12), уникальные в пределах периода — интервала типа *год*.

Интервал типа *квартал* имеет числовой идентификатор (номер от 1 до 4), уникальный в пределах периода — интервала типа *год*.

Интервал типа *полугодие* имеет числовой идентификатор (номер 1 или 2), уникальный в пределах периода — интервала типа *год*.

Интервалы типа *год*, *век* и *тысячелетие* имеют числовой идентификатор — номер, уникальный в пределах одного из двух интервалов, относящихся к типу *эра*):

- (1) интервала с началом летоисчисления в качестве левой границы (значение *н-э*);
- (2) интервала с началом летоисчисления в качестве правой (не включенной) границы (значение *д-н-э*).

Отсутствие уточнения *н. э.* или *до н. э.* при номере года, века и тысячелетия трактуется как значение *н-э*.

Интервал типа *десятилетие* имеет числовой идентификатор (номер от 1 до 10), уникальный в пределах периода — интервала типа *век*.

Интервал типа *день недели* имеет лексический идентификатор (название дня недели), уникальный в пределах периода — интервала типа *неделя*.

Интервал типа *неделя* в языке (в отличие от формального представления времени) не имеет собственных идентификаторов и идентифицируется исключительно средствами дейксиса или анафоры: *на этой (прошлой, той же, будущей) неделе*.

Интервал типа *время года* имеет одно из четырех названий: *ВЕСНА*, *ЛЕТО*, *ОСЕНЬ*, *ЗИМА*. Три названия уникальны в пределах периода — интервала типа *год*, а интервал с названием *ЗИМА*, который приходится на границу между интервалами-годами, идентифицируется в периоде, состоящем из двух лет.

Типы календарных интервалов основной шкалы упорядочены по масштабу следующим образом:

день/сутки < декада < месяц < квартал < полугодие
< год < десятилетие < век < тысячелетие < эра.

Введем понятие *содержащего интервала*.

Для любого единичного календарного интервала a типа A и любого типа B таких, что $A < B$, существует ровно один интервал b типа B , содержащий a . Интервал c типа C , содержащий b , также содержит и a . Любой единичный календарный интервал содержит сам себя.

Что касается интервалов дополнительных шкал, то для них действуют следующие положения:

- (1) день недели $<$ неделя.
- (2) Интервал день недели имеет тот же масштаб, что и день/сутки, при этом любой интервал типа день недели имеет содержащий его интервал день/сутки и наоборот, для любого интервала день/сутки существует содержащий его интервал день недели (и, соответственно, содержащая его неделя).
- (3) Поскольку шкала недель не привязана к основной календарной шкале, то лишь для некоторых интервалов типа неделя верно, что существует содержащий их интервал типа месяц. Для таких недель выполняется и вся транзитивная цепочка (т. е. все интервалы, содержащие месяц, содержат и неделю). Для произвольной недели можно утверждать, что существует содержащий ее календарный интервал некоторого типа³ (если не месяц, то либо квартал, либо полугодие, либо год и т. п.).
- (4) Общим для типа время года является соотношение месяц $<$ время года. Для каждого интервала типа время года, за исключением имеющих значение ЗИМА, существует содержащий его интервал типа год.

Если определен (известен) некоторый интервал i , то известен и единственный интервал любого из типов, имеющих больший масштаб, содержащий (непосредственно или транзитивно) i . Это позволяет при необходимости привести рассматриваемый интервал к интервалу требуемого типа, имеющего больший масштаб. Это обстоятельство будет в дальнейшем использовано при определении понятия интервала отсчета как интервала заданного типа, содержащего время отсчета (раздел 3.1). Поскольку время отсчета — произвольный интервал (не обязательно единичный календарный), сделаем следующую оговорку. Предполагаем, что для произвольного интервала j , не являющегося единичным календарным интервалом, всегда существует содержащий

³Если только эта неделя не располагается по обе стороны от начала летоисчисления — в таком случае ее не содержит ни один интервал определенного в нашей модели типа.

его единичный календарный интервал некоторого типа⁴. Любой интервал, имеющий масштаб меньше суток (таково часто время создания новостного текста), округляется до содержащего его интервала типа *день/сутки*. Считаем, что длительность момента речи (или текущего действительного временного центра) всегда меньше минимальной длительности, различаемой в модели.

1.2. Производные интервалы

В дополнение к уже описанным типам календарных интервалов полезно иметь в модели способ представления некоторых частотных языковых конструкций.

Конструкции типа *последний день (такого-то) месяца, последний квартал (такого-то) года* представляют собой альтернативные способы референции к единичному календарному интервалу и описываются функциями типа: `lastDayOfMonth(monthi)`, `lastQuarterOfYear(yeari)` и т. п.

Вычисление значений этих функций, т. е. собственно определение интервала, не входит в компетенцию лингвистической модели (это задача уровня формальной модели).

В дополнение к единичным календарным интервалам введем интервал типа *выходные*, который представляет собой точный интервал, состоящий из двух конкретных единичных интервалов типа *день недели*: `weekend (weeki)`. При работе с темпоральной информацией полезна возможность оперировать не только точными интервалами, но также интервалами разной степени неточности (неопределенности⁵). Вопросы моделирования НЕ-факторов на уровне лингвистической модели пока не затрагиваются. Ограничимся представлением одного типа интервалов с размытыми границами⁶. Это интервалы, пред-

⁴Если только j не располагается по обе стороны от начала летоисчисления. Но такой интервал маловероятен в роли времени отсчета для контекстной даты.

⁵Мы не беремся здесь давать строгих определений различным *НЕ-факторам* (термин предложен А. С. Нариньяни для обозначения плохо формализуемых знаний, дефектов знания, возможных форм незнания, свойственных любому знанию [13]), поскольку их систематизация применительно к задаче представления времени может составить тему отдельного серьезного исследования.

⁶Здесь речь идет о принципиальной размытости, зафиксированной в модели, а не о неточности в силу несовершенства наших знаний или недостаточной информации. Уточнено может быть время события, локализованного с помощью выражения *в начале этого года*, — например, после уточнения его другим временным указателем. Но точно установить сам интервал, соответствующий выражению *начало этого года*, невозможно в принципе.

ставляющие собой фрагменты (начальный, срединный и конечный) единичного календарного интервала типа X и образованные последовательностью единичных интервалов произвольного типа Y_i , такого что $Y_i < X$). Каждая такая последовательность имеет размытые границы, но может быть ограничена точной последовательностью. В языке для обозначения таких интервалов используются именные группы *начало года, первые дни года, середина года, конец года, начало месяца, середина месяца, конец месяца* и т. п. Для представления таких интервалов используются следующие обозначения: `BeginningOfYear(yeari)`, `FirstDaysOfYear(yeari)`, `MiddleOfYear(yeari)`, `EndOfYear(yeari)` и т. п.

1.3. Расстояние между интервалами

Расстояние между двумя смежными единичными календарными интервалами шкалы S считаем равным единице длительности S . Например, расстояние между интервалами дат *31 декабря 2000 года* и *1 января 2001 года* равно 1 (день/сутки).

1.4. Представление единичного календарного интервала

На типах календарных интервалов определим отношение «идентифицируется в периоде» (PERIOD):

PERIOD = {(день/сутки, месяц); (декада, месяц); (месяц, год); (квартал, год); (полугодие, год); (год, эра); (десятилетие, век); (век, эра); (тысячелетие, эра); (день недели, неделя); (время года⁷; год)}.

Смысл отношения PERIOD в следующем: если a — интервал типа A и PERIOD(A, B), то a однозначно определяется своим собственным идентификатором и содержащим его (интервал a) интервалом-периодом b типа B . Если тип B таков, что PERIOD(B, C), то для определения b , в свою очередь, требуется также и интервал-период c типа C , содержащий b , и т. д.

Определим набор обязательных атрибутов для типа A как кортеж максимальной длины, состоящий из типов интервалов $\langle X_1; X_2; \dots; X_n \rangle$, такой что $X_1 = A$ и для любой пары элементов кортежа $\langle X_i; X_{i+1} \rangle$ имеет место PERIOD(X_i, X_{i+1}). Так, набор обязательных атрибутов для типа *день/сутки* — это кортеж $\langle \text{день/сутки}; \text{месяц}; \text{год}; \text{эра} \rangle$.

⁷Что касается типа *время года*, то будем считать, что интервалы со значениями ВЕСНА, ЛЕТО и ОСЕНЬ идентифицируются в периоде год, а ЗИМА в виде исключения идентифицируется в периоде, состоящем из двух лет.

Таблица 1. Наборы обязательных атрибутов календарных интервалов

Тип интервала	Обозначение единичного интервала	Набор обязательных атрибутов
день/сутки	day	⟨DAY; MONTH; YEAR; ERA⟩
день недели	day_of_week	⟨DAYofWEEK; WEEK⟩
неделя	week	⟨WEEK⟩
декада	ten-days	⟨TEN-DAYS; MONTH; YEAR; ERA⟩
месяц	month	⟨MONTH; YEAR; ERA⟩
время года	season	⟨SEASON; YEAR; ERA⟩
квартал	quarter	⟨QUARTER; YEAR; ERA⟩
полугодие	half-year	⟨HALF-YEAR; YEAR; ERA⟩
год	year	⟨YEAR; ERA⟩
десятилетие	decade	⟨DECADE; CENTURY; ERA⟩
век	century	⟨CENTURY; ERA⟩
тысячелетие	millennium	⟨MILLENNIUM; ERA⟩

Единичный календарный интервал задается в лингвистической модели кортежем пар «атрибут, значение». В качестве имен атрибутов выступают названия типов интервалов, а в качестве значений — идентификаторы. Набор пар в кортеже соответствует набору обязательных атрибутов типа. Для обозначения атрибутов в кортеже используются английские названия соответствующих типов, записанные заглавными буквами, а для обозначения конкретного единичного интервала — англоязычное название его типа, записанное строчными буквами (таблица 1).

Кортеж интервала типа *день/сутки*, соответствующего текстовой дате *1 апреля 2015 года*, выглядит следующим образом: $\text{day}(1 \text{ апреля } 2015 \text{ года}) = \langle \text{DAY} = 1; \text{MONTH} = 4; \text{YEAR} = 2015; \text{ERA} = \text{н-э} \rangle$. Иногда для представления значения контекстных дат вместо развернутой поатрибутной записи будет использоваться частично (полностью) свернутая запись, в которой часть пар (все пары) кортежа вида «атрибут, значение» заменены обозначением единичного интервала, например:

$\langle A = \text{val}_A; B = \text{val}_B; C = \text{val}_C \rangle$ — развернутая поатрибутная запись интервала;

$\langle A = \text{val}_A; b \rangle$ — частично свернутая запись интервала;

a — полностью свернутая запись интервала.

2. Структура и виды текстовых дат

Текстовая дата — текстовое выражение (именная группа, буквенно-цифровое обозначение) с референцией к конкретному единичному календарному интервалу⁸ [1]. Текстовая дата может быть абсолютной (если содержит все обязательные атрибуты интервала в явном виде) или контекстной (если вместо части значений содержит индексальные показатели или эллипсис).

Будем говорить о типе текстовой даты, понимая под этим тип ее календарного интервала (например, *прошлый понедельник* — это текстовая дата типа *день недели*).

Как правило, в поверхностной структуре текстовой даты (далее ТД) можно выделить компоненты, соответствующие обязательным атрибутам ее календарного интервала, — назовем эти компоненты полями ТД.

Первое поле — собственное поле ТД — всегда соответствует первому атрибуту в кортеже, определяющему тип интервала ТД. Остальные поля — поля периодов (поля интервалов-периодов, содержащих интервал ТД). Собственное поле является обязательным полем любой ТД; поля периодов в поверхностной структуре ТД могут отсутствовать.

Абсолютная дата всегда содержит полный⁹ набор полей; в каждом поле присутствует явное указание на значение соответствующего атрибута.

В контекстной дате может иметь место что-либо из следующего:

- (1) в одном или нескольких полях ТД содержится отсылочный показатель (анафорический или дейктический компонент — *тот, этот, тот же, этот же, прошлый, позапрошлый, предшествующий, настоящий, текущий, будущий, следующий*);
- (2) в составе ТД нет отсылочных показателей и при этом (далее через «и/или»):

⁸Выражения с референцией к интервалам производных типов не относятся к текстовым датам. Их значения могут быть описаны функциями, аргументами которых выступают интервалы текстовых дат.

⁹В отсутствие поля *эра* ему приписывается значение н-э.

- отсутствует одно или несколько полей периодов¹⁰, например *1-е число* (отсутствуют поля периодов *месяц* и *год*); *апрель, второй квартал, лето* [*летом*] (эллипсис поля *год*), [*в*] *понедельник* (эллипсис поля *неделя*);
- в поле периода отсутствует указание на значение атрибута, например: *1-е число* \emptyset *месяца* (эллипсис значения в поле *месяц*, эллипсис поля *год*), *второй квартал* \emptyset *года* (эллипсис значения в поле *год*), [*на*] \emptyset *неделе* (эллипсис значения в поле *неделя*).

ТД первой группы будем называть *ТД с явными отсылочными показателями*. ТД второй группы — *эллиптическими ТД*.

Заметим, что ТД с отсутствующим полем не считается эллиптической, если последнее из имеющихся полей содержит отсылочный показатель. Так, в ТД *15 числа прошлого месяца* нет эллипсиса поля *год*, т. к. его значение «включено» в значение *месяца* и автоматически восстановится при вычислении последнего.

ТД *прошлый понедельник, будущий сентябрь, это лето* и т. п., состоящие из названия (лексического идентификатора) интервала типа *день недели, месяц или время года* и индексального показателя, рассматриваем как отдельный тип ТД, без выделения в них полей (в отличие от ТД *понедельник на прошлой неделе, сентябрь будущего года и лето этого года*, где выделяются собственное поле и поле периода).

Отдельно отметим случаи, когда для референции к единичному календарному интервалу используются два самостоятельных адвербиальных выражения (одно указывает на значение собственного интервала, другое — на значение интервала-периода), например:

- в 2001 году, в мае;*
- в прошлом году, 15 мая;*
- в мае, 15-го (числа);*
- два года назад, в мае;*
- на прошлой неделе, в понедельник*¹¹;
- в этот день три года назад;*
- в прошлом году в этот день.*

¹⁰За исключением поля *эра*: его отсутствие не считаем эллипсисом, подлежащим восстановлению (по умолчанию атрибут ERA заполняется значением н-э).

¹¹Выражение *понедельник на прошлой неделе* считаем стандартной текстовой датой с отсылочным показателем в поле *неделя*.

Подобные сочетания упоминаний могут быть кореферентны стандартным ТД, как абсолютным, так и контекстным. Такие случаи в настоящей работе не рассматриваются, заметим только, что их семантическое значение представимо описываемыми здесь средствами.

3. Контекстные даты с явными отсылочными показателями

3.1. Интервал отсчета

Отсылочный показатель в составе ТД содержит в себе анафорическую или дейктическую ссылку на некоторое фиксированное время, относительно которого и определяется (задается) значение интервала ТД (собственного или периода) или ее отдельного атрибута. Интервал, к которому ведет такая ссылка, будем называть *временем отсчета* для отсылочного показателя ТД.

Новостной текст является разновидностью нарратива 3 лица. В качестве времени отсчета для отсылочного показателя в нем может выступать:

- *время создания текста* (глобальный дейктический временной центр, аналог момента речи в канонической речевой ситуации) — при нарративной¹² интерпретации дейктических отсылочных показателей;
- *референциальный интервал* упомянутого в самом тексте показателя времени (временного antecedента) или время упомянутого в тексте события — для анафорических отсылок, а также при персональной интерпретации дейктических показателей (в последнем случае время отсчета — это текущий дейктический центр¹³, момент речи внутритекстового наблюдателя).

Разрешение дейктической или анафорической ссылки в составе контекстной даты, т.е. собственно определение того, что является временем отсчета для отсылочного показателя, — это самостоятельная задача, которая выходит за рамки настоящей работы. Предлагаемый далее способ представления значения текстовых дат с отсылочными показателями предполагает, что время отсчета может быть установлено.

¹²Нарративная интерпретация эгоцентрических элементов ориентирована на повествователя, а персональная — на персонаж [14].

¹³Далее для простоты будем использовать термин *момент речи* для обозначения времени отсчета при любой дейктической отсылке, т.е. времени создания текста или текущего дейктического центра повествования.

Для представления значения ТД с отсылочными показателями введем понятие интервала отсчета.

Пусть r_0 — время отсчета для отсылочного показателя в поле A контекстной ТД (A — тип календарного интервала). Тогда интервалом отсчета для этого отсылочного показателя назовем единичный календарный интервал типа A , содержащий время отсчета¹⁴.

Например:

«В позапрошлом году я побывал на Камчатке».

Интервалом отсчета для отсылочного показателя позапрошлый в данном случае является *год отсчета* $year_0$, содержащий время отсчета r_0 (в данном случае r_0 — момент речи).

Заметим, что отсылочные показатели в разных полях одной и той же ТД могут содержать отсылку к разному времени отсчета (соответственно, интервалы отсчета тоже будут разные), например:

«В том же квартале прошлого года результаты были значительно хуже.»

Тот же квартал — анафорическая отсылка к времени отсчета, упомянутому в предшествующем фрагменте; *прошлый год* — дейктическая отсылка к времени создания текста.

Идентификаторы (значения) атрибутов интервала отсчета будем обозначать как DAY_0 , $MONTH_0$, $YEAR_0$ и т. п.

3.2. Интерпретация отсылочных показателей в составе ТД. Функции

В семантическом значении отсылочного показателя, входящего в состав поля ТД, можно выделить две составляющие. Одна — *индексальное значение* — указывает на характер отсылки (дейктический или анафорический) к времени отсчета. Вторая — назовем ее *операциональным значением* — это по существу инструкция, которую необходимо выполнить адресату для получения значения атрибута или интервала ТД по интервалу отсчета.

Операциональное значение отсылочного показателя представим в виде функции одного из трех типов (подробнее описанных далее):

¹⁴Предполагаем, что если известно время отсчета для отсылочного показателя в поле A , то известен и интервал масштаба A , содержащий время отсчета. Хотя в художественном повествовании возможны ситуации, когда это не так, например: *Это произошло в 1999 году. В тот день...* [в день, когда это произошло, — но точная дата типа *день/сутки* не упоминается].

- (1) функция заимствования значения атрибута или интервала без модификации;
- (2) функция заимствования и модификации значения атрибута или интервала;
- (3) функция поиска ближайшего интервала с заданным идентификатором.

Значение поля ТД, содержащего отсылочный показатель, может быть описано в виде *формулы* с оператором присваивания « := ». В левой части формулы — интервал (имя атрибута), которому присваивается значение. В правой части — выражение, содержащее некоторую функцию. В роли переменной выступает интервал отсчета (кортеж его значений) или значение его отдельного атрибута. Каждое такое выражение должно быть однозначно «понято» формальной моделью, которая выполнит подстановку значений в переменные и вычислит результат.

3.2.1. Функция заимствования значения без модификации

Функция заимствует у интервала отсчета значение отдельного атрибута или кортеж значений интервала.

Формула для вычисления значения отдельного атрибута ТД:

$$A := A_0,$$

где A — имя атрибута ТД, A_0 — значение атрибута типа A интервала отсчета.

Например, в ТД *тот же квартал прошлого года* отсылочный показатель *тот же* в поле *квартал* указывает на то, что значение соответствующего атрибута (и только оно) должно быть заимствовано у интервала отсчета; остальные атрибуты в кортеже ТД должны быть определены независимо от него.

Формула для вычисления значения интервала (самой ТД или ее периода):

$$a := a_0,$$

где a — собственный интервал ТД или один из содержащих его интервалов-периодов, a_0 — интервал отсчета.

Формулу $a := a_0$ можно представить как поатрибутное присвоение:

$$a = \langle A := A_0; B := B_0; \dots \rangle$$

Например, ТД в составе наречия *сегодня* может быть представлена формулой:

$$\text{day} = \langle \text{DAY} := \text{DAY}_0; \text{MONTH} := \text{MONTH}_0; \text{YEAR} := \text{YEAR}_0; \text{ERA} := \text{ERA}_0 \rangle,$$

где DAY_0 , MONTH_0 , YEAR_0 , ERA_0 — значения соответствующих атрибутов дня отсчета.

3.2.2. Функция заимствования значения с модификацией

Функция заимствует у интервала отсчета значение отдельного атрибута (кортеж значений интервала) и выполняет над ним операцию модификации.

3.2.2.2. Заимствование значения атрибута с модификацией

Операция *модификации значения отдельного атрибута* $X \pm n$, где X — исходное значение (идентификатор) атрибута типа S , n — число единиц длительности S , прибавляемое к значению атрибута S или отнимаемое от него. Результатом выполнения операции (в формальной модели) является идентификатор интервала типа S , отстоящего от календарного интервала с идентификатором X на n единиц длительности S . Так, если $S = \text{месяц}$ и $X = \text{СЕНТЯБРЬ}$, то $X - 1 = \text{АВГУСТ}$; если $S = \text{квартал}$ и $X = 1$, то $X + 1 = 2$ и т. п. Операция модификации значения отдельного атрибута определена не для всех значений X и n , а только для таких, при которых интервал, отстоящий от интервала с идентификатором X на n единиц длительности, находится в том же периоде. Так, для типа *квартал* операция $X - 1$ определена только для значений 2, 3 и 4, операция $X + 1$ — для значений 1, 2 и 3.

Формула для вычисления значения *отдельного атрибута* кортежа ТД:

$$A := A_0 \pm n,$$

где A_0 — значение атрибута A интервала отсчета, а n — число единиц длительности A .

Так, в ТД *предшествующий квартал 2012 года* отсылочный показатель *предшествующий* предписывает применить к значению атрибута QUARTER квартала отсчета (и только к нему, поскольку интервал периода-года в ТД фиксирован) операцию «минус один квартал».

3.2.2.2. Заимствование значения интервала с модификацией

Операция *модификации интервала* $x \pm n$, где x — интервал типа S , n — число единиц длительности S , прибавляемое к x или отнимаемое от него. Результатом операции является интервал (кортеж значений интервала) типа S , отстоящий от x на n единиц длительности S .

Формула для вычисления значения *интервала*:

$$a := a_0 \pm n,$$

где a — собственный интервал ТД или интервал-период, a_0 — интервал отсчета, n — число единиц длительности A (A — тип интервалов a и a_0).

Значение интервала, получаемого в результате операции модификации, невозможно представить поатрибутно, в виде формул для значений всех атрибутов.

Приведем пример: предположим, время отсчета для отсылочного показателя, «встроенного» в поле *день/сутки* ТД *вчера*, приходится на 1 января 2015 года, т.е. день отсчета для отсылочного показателя описывается кортежем: $day_0 = \langle DAY_0 = 1; MONTH_0 = 1; YEAR_0 = 2015; ERA_0 = н-э \rangle$. Значение самой ТД в составе наречия *вчера* описывается формулой:

$$day := day_0 - 1,$$

что в данном случае будет соответствовать интервалу: $\langle DAY = 31; MONTH = 12; YEAR = 2014; ERA = н-э \rangle$.

Таким образом, вычитание одного дня из значения конкретного интервала типа *день/сутки* может привести к изменению не только значения собственного атрибута (DAY), но также значений месяца и года. А если бы время отсчета приходилось на 2 января, то атрибуты $MONTH$ и $YEAR$ сохранили бы значения, заимствованные у интервала отсчета.

3.2.3. Функция поиска ближайшего интервала с заданным идентификатором.

$NEAREST_AFTER(X, a_0)$, где X — идентификатор интервала типа A , $A = \{\text{день недели, месяц, время года}\}$, a_0 — интервал типа A , содержащий время отсчета.

$NEAREST_BEFORE(X, a_0)$, где X — идентификатор интервала типа A , $A = \{\text{день недели, месяц, время года}\}$, a_0 — интервал типа A , содержащий время отсчета.

Значением функции является ближайший к времени отсчета (после или до него, соответственно) интервал заданного типа, имеющий заданный идентификатор. Время отсчета — момент речи.

Например, значение ТД *ближайшее воскресенье* описывается формулой:

$$day_of_week := NEAREST_AFTER(ВОСКРЕСЕНЬЕ, day_of_week_0).$$

3.3. Виды отсылочных показателей по функции

В таблице 2 отсылочные показатели сгруппированы по операциональному значению, которым они обладают в составе той или иной ТД:

показатели заимствования — показатели, которые интерпретируются как функции заимствования без модификации;

показатели модификации — показатели, которые интерпретируются как функции заимствования с модификацией; внутри эта группа подразделяется на подгруппы по знаку операции и значению параметра n ;

показатели поиска ближайшего — показатели, которые интерпретируются как функции поиска ближайшего интервала с заданным идентификатором.

Комментарии к таблице. Анафорическое (катафорическое) *тот* относится к показателям заимствования, тогда как его дейктические омонимы (свойственные скорее разговорной речи) — к показателям модификации. Дейктическое *тот* может иметь значение 'прошлый', а иногда и 'следующий': *на той* ('прошлой' или 'будущей') *неделе*, *в том* ('прошлом') *году*, *тем* ('прошлым') *летом*.

Заметим, что оба омонима в парах «*этот* (анафорическое) — *этот* (дейктическое)», «*этот же* (анафорическое) — *этот же* (дейктическое с усилительной частицей)» имеют одинаковое операциональное значение, но разное индексальное (оба относятся к показателям заимствования, но содержат отсылки к времени отсчета, имеющие разную природу).

Как видно из таблицы, группа показателей поиска ближайшего целиком образована из тех же прилагательных, которые входят в группы модификации «+1» и «-1». Как именно интерпретируется показатель в составе конкретной ТД, зависит от семантики определяемого им слова (тип интервала или идентификатор интервала типа *день недели*, *месяц*, *время года*).

Прилагательному *последующий* не свойственно употребление с референцией к конкретному единичному календарному интервалу (например, *каждый последующий год* (*квартал*, *месяц*, *день*) — выражения с квантором общности). Как ТД может употребляться лишь выражение *последующие сутки*, хотя и оно может обозначать небольшой промежуток длиной в несколько суток.

Таблица 2. Отсылочные показатели в составе текстовых дат

Показатели заимствования	
	<i>этот, тот</i> (анафорич.), <i>тот самый, этот же, тот же, текущий, нынешний, сегодня, наш</i> (с существительными <i>век, тысячелетие, эра</i>)
Показатели модификации	
-2	<i>позапрошлый, позавчера</i> <i>прошлый, предшествующий, предыдущий, прошедший</i> (с указанием на тип интервала), -1 <i>минувший</i> (аналогично), <i>вчера, накануне, тот</i> (дейктич., в значении 'прошлый') <i>следующий, последующий</i> (с сущ. <i>сутки</i>), <i>тот</i> (дейктич., в значении 'следующий'), <i>завтра</i> , +1 <i>наступающий</i> (с указанием на тип интервала), <i>будущий</i> (аналогично), <i>ближайший</i> (аналогично), <i>предстоящий</i> (аналогично), <i>назавтра</i> +2 <i>послезавтра</i>
Показатели поиска <i>ближайшего X</i>	
до	<i>минувший</i> (с названием интервала типа <i>день недели, месяц, время года</i>), <i>прошедший</i> (аналогично) <i>будущий</i> (с названием интервала типа <i>день недели, месяц, время года</i>), <i>ближайший</i> (аналогично), после <i>наступающий</i> (аналогично), <i>предстоящий</i> (аналогично)

Считаем, что в составе наречий *сегодня, завтра, назавтра, вчера, позавчера, накануне* подлежит реконструкции текстовая дата типа *день/сутки* с встроенным отсылочным показателем: заимствования (*сегодня* = 'в сегодняшшний день', в день, совпадающий с днем момента речи) или модификации (во всех остальных случаях).

3.4. Представление значений полей ТД, содержащих отсылочные показатели

3.4.1. Поле вида «отсылочный показатель + имя типа интервала»

Пусть Тип(*A*) — лексема, обозначающая тип интервала *A* (*день, сутки* — *день/сутки, век, столетие* — *век* и т. п.), ОП — отсылочный

Таблица 3. Показатель заимствования при лексеме со значением типа интервала

Тип	Закрытая позиция	Открытая позиция (поатрибутная запись интервала)
день/сутки	$DAY = DAY_0$	$day = \langle DAY := DAY_0;$ $MONTH := MONTH_0;$ $YEAR := YEAR_0; ERA := ERA_0 \rangle$
неделя	невозможна (нет периода)	$week = \langle WEEK := WEEK_0 \rangle$
декада	$TEN-DAYS$ $= TEN-DAYS_0$	$ten-days = \langle TEN-DAYS := TEN-DAYS_0;$ $MONTH := MONTH_0;$ $YEAR := YEAR_0; ERA := ERA_0 \rangle$
месяц	$MONTH = MONTH_0$	$month = \langle MONTH := MONTH_0;$ $YEAR := YEAR_0; ERA := ERA_0 \rangle$
квартал	$QUARTER$ $= QUARTER_0$	$quarter = \langle QUARTER := QUARTER_0;$ $YEAR := YEAR_0; ERA := ERA_0 \rangle$
полугодие	$HALF-YEAR$ $= HALF-YEAR_0$	$half-year = \langle HALF-YEAR := HALF-YEAR_0;$ $YEAR := YEAR_0; ERA := ERA_0 \rangle$
год	$YEAR = YEAR_0$	$year = \langle YEAR := YEAR_0; ERA := ERA_0 \rangle$
десятилетие	$DECADE$ $= DECADE_0$	$decade = \langle DECADE := DECADE_0;$ $CENTURY = CENTURY_0; ERA := ERA_0 \rangle$
век	$CENTURY$ $= CENTURY_0$	$century = \langle$ $CENTURY := CENTURY_0; ERA := ERA_0 \rangle$
тысячелетие	$MILLENNIUM$ $= MILLENNIUM_0$	$millennium = \langle$ $MILLENNIUM := MILLENNIUM_0;$ $ERA := ERA_0 \rangle$

показатель. Интерпретация ОП зависит от того, какую позицию поле «ОП + Тип(A)» занимает в ТД.

3.4.1.1. Закрытая позиция

Закрытой называем такую позицию поля «ОП + Тип(A)», когда за этим полем в ТД следует поле периода типа A (поле, соответствующее периоду типа A). В этом случае сфера действия ОП ограничена атрибутом соответствующего A типа. Это значит, что в кортеже ТД поле, содержащее ОП, будет представлено формулой для отдельного атрибута (таблица 3, столбец «Закрытая позиция»).

Закрытая позиция характерна главным образом для ОП — показателей заимствования: *того же числа прошлого месяца, в той же*

ТАБЛИЦА 4. Показатель модификации при лексеме со значением типа интервала

Тип	Закрытая позиция	Открытая позиция
день/сутки	$DAY := DAY_0 \pm n$	$day := day_0 \pm n$
неделя	невозможна (нет периода)	$week = week_0 \pm n$
декада	$TEN-DAYS := TEN-DAYS_0 \pm n$	$ten-days := ten-days_0 \pm n$
месяц	$MONTH := MONTH_0 \pm n$	$month := month_0 \pm n$
квартал	$QUARTER := QUARTER_0 \pm n$	$quarter := quarter_0 \pm n$
полугодие	$HALF-YEAR := HALF-YEAR_0 \pm n$	$half-year$ $:= half-year_0 \pm n$
год	$YEAR := YEAR_0 \pm n$	$year = year_0 \pm n$
десятилетие	$DECADE := DECADE_0 \pm n$	$decade := decade_0 \pm n$
век	$CENTURY := CENTURY_0 \pm n$	$century := century_0 \pm n$
тысячелетие	$MILLENNIUM$ $:= MILLENNIUM_0 \pm n$	$millennium$ $:= millennium_0 \pm n$

декаде мая, в этом же квартале будущего года.

В кортеже ТД атрибуту A присваивается значение атрибута A интервала отсчета. Значения других атрибутов определяются независимо.

Показателям модификации закрытая позиция не слишком свойственна, хотя в реальных текстах встречаются ТД вроде *предшествующий квартал 2010 года, прошлый месяц этого (того же) года.*

В кортеже такой ТД атрибуту типа A приписывается выражение с той операцией модификации, на которую указывает ОП (таблица 4, столбец «Закрытая позиция»).

Возможен вариант закрытой позиции, когда за полем «ОП + Тип(A)» следует поле интервала, не являющегося собственным периодом A . Практически эта возможность реализуется только в ТД типа *день/сутки*: встречается, например, выражение *в этот же (тот же) день прошлого года* (хотя более естественным представляется *в этот (же) день в прошлом году* или *год назад*). В этом случае следует считать, что ТД включает поля *день/сутки* и *месяц* с показателем тождества в закрытой позиции.

3.4.1.2. Открытая позиция

Открытой назовем конечную позицию поля «ОП + Тип(A)» в ТД (при этом A может и не быть последним из обязательных атрибутов

Таблица 5. Примеры значений контекстных дат

Контекстная дата	Значение
<i>то же число прошлого месяца</i>	$\text{day} = \langle \text{DAY} := \text{DAY}_0; \text{month} := \text{month}_0 - 1 \rangle$
<i>тот же день прошлого года (в прошлом году)</i>	$\text{day} = \langle \text{DAY} := \text{DAY}_0; \text{MONTH} = \text{MONTH}_0;$ $\text{year} = \text{year}_0 - 1 \rangle$
<i>та же декада будущего месяца</i>	$\text{ten-days} = \langle \text{TEN-DAYS} := \text{TEN-DAYS}_0;$ $\text{month} := \text{month}_0 + 1 \rangle$
<i>предшествующий квартал текущего года</i>	$\text{quarter} = \langle \text{QUARTER} := \text{QUARTER}_0 - 1;$ $\text{YEAR} = \text{YEAR}_0; \text{ERA} = \text{ERA}_0 \rangle$
<i>понедельник на этой неделе</i>	$\text{day_of_week} =$ $\langle \text{DAY_of_WEEK} := \text{ПОНЕДЕЛЬНИК};$ $\text{WEEK} := \text{WEEK}_0 \rangle$
<i>первая декада прошлого месяца</i>	$\text{ten-days} = \langle \text{TEN-DAYS} := 1;$ $\text{month} := \text{month}_0 - 1 \rangle$
<i>9 мая позапрошлого года</i>	$\text{day} = \langle \text{DAY} = 9; \text{MONTH} = \text{МАЙ};$ $\text{year} := \text{year}_0 - 2 \rangle$
<i>предшествующее полугодие</i>	$\text{half-year} := \text{half-year}_0 - 1$
<i>нынешнее тысячелетие</i>	$\text{millennium} = \langle \text{MILLENNIUM} := \text{MILLENNIUM}_0;$ $\text{ERA} := \text{ERA}_0 \rangle$

типа). Частный случай открытой позиции имеет место, если поле «ОП + Тип(A)» является единственным полем ТД.

В открытой позиции встречаются как показатели заимствования, так и показатели модификации.

Действие ОП в таких случаях распространяется не на один атрибут ТД, а на интервал типа A , содержащий интервал ТД (в случае единственного поля в ТД — на собственный интервал ТД). Соответственно, в кортеже ТД такое поле представляется формулой для значения интервала (см. столбцы «Открытая позиция» в таблицах 3 и 4).

Примеры: *на этой неделе, вчера, в первой декаде прошлого месяца, 9 мая позапрошлого года, в предыдущем квартале, в предшествующем полугодии, в нынешнем тысячелетии.*

Примеры того, как представляются значения контекстных ТД, содержащих поле с ОП в различных позициях, приведены в таблице 5.

Кортежи ТД типа *день/сутки*, реконструируемых в дейктических

ТАБЛИЦА 6. Текстовые даты в составе наречий

Наречие	Значение
<i>позавчера</i>	$\text{day} := \text{day}_0 - 2$
<i>вчера, накануне</i>	$\text{day} := \text{day}_0 - 1$
<i>сегодня</i>	$\text{day} = \langle \text{DAY} := \text{DAY}_0; \text{MONTH} := \text{MONTH}_0; \text{YEAR} := \text{YEAR}_0; \text{ERA} := \text{ERA}_0 \rangle$
<i>завтра, наавтра</i>	$\text{day} := \text{day}_0 + 1$
<i>послезавтра</i>	$\text{day} := \text{day}_0 + 2$

Таблица 7. Показатель модификации при названии интервала

X — идентификатор интервала типа:	Значение ТД вида «показатель модификации + X »
день недели	$\text{day_of_week} = \langle \text{DAY_of_WEEK} := X; \text{WEEK} := \text{WEEK}_0 \pm n \rangle$
месяц	$\text{month} = \langle \text{MONTH} := X; \text{year} := \text{year}_0 \pm n \rangle$
время года	$\text{season} = \langle \text{SEASON} := X; \text{year} := \text{year}_0 \pm n \rangle$

и анафорических наречиях, приведены в таблице 6.

3.4.2. ТД вида «отсылочный показатель + название интервала»

3.4.2.1. Показатель модификации + название интервала

Отсылочный показатель модификации (за исключением тех, которые являются также и показателями поиска ближайшего) при названии интервала типа *день недели*, *месяц*, *время года* интерпретируется как значение интервала-периода, например: *следующий понедельник* = 'понедельник на следующей неделе', *позапрошлый вторник* = 'вторник на позапрошлой неделе', *будущий сентябрь* = 'сентябрь в будущем году', *прошлое лето* = 'лето прошлого года'. Коротежи с формулами для ТД такого вида приведены в таблице 7.

3.4.2.2. Показатель заимствования + название интервала

В общем случае значение ТД, состоящей из отсылочного показателя заимствования и названия (лексического идентификатора) интервала типа *день недели*, *месяц*, *время года*, описывается формулой,

Таблица 8. Показатель заимствования при названии интервала

X — идентификатор интервала типа:	Значение ТД вида «показатель заимствования + X »
день недели	$\text{day_of_week} = \langle \text{DAY_of_WEEK} := X; \text{WEEK} := \text{WEEK}_0 \rangle$
месяц	$\text{month} = \langle \text{MONTH} := X; \text{YEAR} := \text{YEAR}_0; \text{ERA} := \text{ERA}_0 \rangle$
время года	$\text{season} = \langle \text{SEASON} := X; \text{YEAR} := \text{YEAR}_0; \text{ERA} := \text{ERA}_0 \rangle$

в которой в значение собственного атрибута подставляется идентификатор из ТД, а значение периода заимствуется из интервала отсчета (таблица 8).

Таким образом, *этот четверг* трактуется как 'четверг на этой неделе', *этот сентябрь* — 'сентябрь этого года', *этой весной* — 'весной этого года' и т. п.

Заметим, что дейктический показатель заимствования в таких ТД указывает на совпадение значения периода ТД с периодом отсчета, тогда как анафорический/катафорический показатель заимствования — скорее, на совпадение самого интервала ТД с интервалом отсчета (в качестве интервала отсчета тогда выступает не период, а конкретный интервал, например *та* (анафорич., *та самая*) *пятница* — это не 'пятница на той самой неделе', а именно 'та конкретная пятница'¹⁵). Однако на уровне формул эти различия не проявляются.

Сделаем еще одно замечание. Дейктическое *этот* в сочетании с названием дня недели может означать не 'находящийся в том же периоде, что и момент речи', а 'ближайший к моменту речи' (независимо от периода), причем как предшествующий моменту речи (*в это воскресенье состоялось*), так и находящийся в будущем (*в этот понедельник состоится*). Такое употребление кажется естественным в ситуации, когда один из двух одноименных интервалов (относящихся к разным смежным периодам) располагается на оси времени явно ближе к моменту речи, чем другой.

¹⁵В новостных текстах иногда встречается аналогичное употребление дейктического *этот* в сочетании с названием дня недели — в значении 'сегодня', например: *Избирком Мариуполя отказался проводить выборы в это воскресенье* (дата создания текста — 25 октября 2015 года, воскресенье). Представление таких ТД на общих основаниях вполне корректно с точки зрения результата.

Таблица 9. Примеры ТД с показателем поиска ближайшего

Контекстная дата	Значение
<i>будущее воскресенье</i>	day_of_week := NEAREST_AFTER (ВОСКРЕСЕНЬЕ, day_of_week ₀)
<i>ближайший сентябрь</i>	month := NEAREST_AFTER(СЕНТЯБРЬ, month ₀)
<i>минувшая зима</i>	season := NEAREST_AFTER (ЗИМА, season ₀)

Например, в *это воскресенье* и даже в *эту субботу*, сказанные в понедельник, вполне могут относиться к прошедшим выходным (т.е. находящимся в предшествующем моменту речи периоде). В общем случае средствами лингвистического анализа отличить *этом* 'ближайший' от *этом* 'находящийся в текущем периоде' затруднительно, если вообще возможно. Не исключено, что «*этом* + название дня недели» следует всегда интерпретировать как 'ближайший к моменту речи, удовлетворяющий ситуационному контексту', независимо от периода, — аналогично случаям ситуативного эллипсиса значения недели (см. об этом в разделе 4.4.2).

3.4.2.3. Показатель поиска ближайшего + название интервала

Значение ТД, представляющих собой сочетание отсылочного показателя *будущий, ближайший, наступающий* и названия интервала типа *день недели, месяц, время года*, описывается формулой, содержащей в левой части обозначение интервала соответствующего ТД типа, а правой части — выражение с функцией NEAREST_AFTER(X, a_0), где X — идентификатор интервала типа A ($A = \{\text{день недели, месяц, время года}\}$), a_0 — интервал типа A , содержащий время отсчета (момент речи).

Для описания аналогичных ТД с показателями *прошедший, минувший* используется функция NEAREST_BEFORE(X, a_0).

Примеры приведены в таблице 9.

4. Эллиптические текстовые даты

Различают два вида эллипсиса: анафорический и ситуативный [15]. В случае анафорического эллипсиса опущенный элемент может быть восстановлен посредством обращения к текстуальному контексту

(обычно — предшествующему фрагменту), при ситуативном эллипсисе — из контекста ситуационного: либо внешнего по отношению к тексту прагматического контекста, либо из ситуации, представляемой в самом тексте.

4.1. Эллипсис указания на тип интервала при наличии значения интервала

В ТД опущено существительное со значением типа интервала (в поле собственного интервала или в поле периода), но имеется идентификатор этого интервала или отсылочный показатель¹⁶, например: *в первом квартале это составило x процентов, а в следующем 0 — y.*

Это однозначно анафорический тип эллипсиса. Для восстановления опущенного элемента в поверхностной структуре необходимо выполнить некоторые синтаксические трансформации. Полученная в результате этого ТД должна быть обработана соответствующим её типу образом (как абсолютная или контекстная ТД).

4.2. Эллипсис значения в поле периода ТД

Примеры: *в первой декаде 0 месяца, во втором квартале 0 года.*

Эллипсис такого типа может быть как анафорическим, так и ситуативным. Механизм восстановления опущенного элемента не зависит от природы эллипсиса: ТД с эллипсисом значения в поле периода может быть уподоблена неэллиптической ТД, содержащей в соответствующем поле отсылочный показатель заимствования — *тот же* (в случае анафорического эллипсиса) или *этот, текущий* (в случае эллипсиса ситуативного), и описана соответствующей формулой. Значение отсутствующего периода заимствуется у интервала отсчета, в роли которого выступает либо интервал ТД-антецедента, либо интервал заданного типа, содержащий время создания текста. Примеры приводятся в таблице 10.

4.3. Эллипсис значения в собственном поле ТД

Этот тип эллипсиса в текстовых датах не слишком распространен. Ситуативный эллипсис значения недели возможен в сочетании *на неделе*. Опущенное значение следует заимствовать у недели, содержащей момент речи (*на неделе* = 'на этой неделе').

¹⁶Ср. эллипсис с сохранением представителя [15].

Таблица 10. Эллипсис значения в поле периода — примеры

<i>первая декада</i> ∅ <i>месяца</i> (того же или текущего месяца)	\langle DECADE := 1; MONTH := MONTH ₀ ; YEAR := YEAR ₀ ; ERA := ERA ₀ \rangle
<i>первое полугодие</i> ∅ <i>года</i> (того же или текущего года)	\langle HALF-YEAR := 1; YEAR := YEAR ₀ ; ERA := ERA ₀ \rangle
<i>второй квартал</i> ∅ <i>года</i> (того же или текущего года)	\langle QUARTER := 2; YEAR := YEAR ₀ ; ERA := ERA ₀ \rangle

Таблица 11. Эллипсис поля периода

<i>5-го числа</i> ∅ (того же или текущего месяца)	\langle DAY := 5; MONTH := MONTH ₀ ; YEAR := YEAR ₀ ; ERA := ERA ₀ \rangle
<i>второй квартал</i> ∅ (того же или текущего года)	\langle QUARTER := 2; YEAR := YEAR ₀ ; ERA := ERA ₀ \rangle
<i>первое полугодие</i> ∅ (того же или текущего года)	\langle HALF-YEAR := 1; YEAR := YEAR ₀ ; ERA := ERA ₀ \rangle

4.4. Эллипсис поля периода

Эллипсис такого типа также может быть как анафорическим, так и ситуативным. При этом имеет смысл различать две группы случаев.

4.4.1. Эллипсис поля периода после поля вида «тип интервала + значение интервала»

Примеры: *в первой декаде, в третьем квартале, во втором полугодии.*

Эти случаи аналогичны уже описанному эллипсису значения в поле периода: эллиптическая ТД может быть уподоблена неэллиптической с анафорическим или дейктическим показателем заимствования в поле периода (таблица 11).

Эллиптические ТД типа *день/сутки* вида *такого-то числа* также с большой вероятностью могут быть отнесены к этой категории.

4.4.2. Эллипсис поля периода после поля, содержащего название интервала

Сюда же относятся ТД, состоящие из одного названия интервала типа *день недели, месяц, время года* (т. е. не содержащие отсылочных показателей).

Примеры: *во вторник, 5 марта, в августе, весной*. В таком контексте разница между анафорическим и ситуативным эллипсисом может иметь принципиальное значение для выбора способа восстановления опущенного элемента.

Если эллипсис анафорический, то оправдана уже описанная тактика заимствования значения эллиптического периода у интервала ТД-антецедента. Например:

«*Во вторник на прошлой неделе состоялось заседание совета директоров. В среду [очевидно, на той же неделе] стало известно. . .*»

«*Изделия, выпущенные прошлым летом и осенью [тоже прошлой]...*»

Если же эллипсис поля периода в поле идентификатора интервала не анафорический, а ситуативный, то подстановка текущего периода (заимствованного у интервала отсчета) может привести к ошибочному результату:

«*В Тверском районном суде Москвы проходит рассмотрение дел участников акции в поддержку Алексея Навального. Мероприятие прошло 30 декабря на Манежной площади.*» Дата создания текста — 12.01.2015.

«*Сегодня в Дохе открывается заседание Форума стран-экспортеров газа (ФСЭГ), которое должно избрать генсека этой организации. По данным «Ъ», им должен стать россиянин — в декабре Москва согласилась на расположение штаб-квартиры организации в Дохе в обмен на право выдвинуть своего генсека.*» Дата создания текста — 30.06.2009.

«*В воскресенье состоялся массовый велопробег „История в движении“.*» Дата публикации — 7 июля 2015 года — приходится на вторник.

«*В пятницу на турбазе было достаточно немногочленно.*» Время создания текста 29 июня 2009 года — понедельник.

Во всех примерах имеет место ситуативный эллипсис, и во всех случаях опущенный период эллиптической ТД не совпадает с периодом, содержащим момент речи (с годом создания текста в первых двух примерах и неделях — в третьем).

Не исключено, что в такой ситуации задача *восстановления эллиптического периода* (путем заимствования периода интервала отсчета, содержащего момент речи) должна быть замещена задачей поиска *ближайшего интервала* с таким названием (до или после момента речи). Представление значения контекстной даты в таком случае выходит за пределы семантического анализа, поскольку даже выбор типа

функции для восстановления эллипсиса (заимствование периода или поиск ближайшего интервала) зависит от типа эллипсиса и потому требует анализа дискурса. Кроме того, если диагностирован ситуативный эллипсис, то для выбора направления поиска ближайшего (функции `NEAREST_AFTER` или `NEAREST_BEFORE`), необходимо установить временные отношения между моментом речи и временем события, на которое указывает эллиптическая дата (точнее, содержащее ее указание на время), а это далеко не тривиальная задача лингвистического анализа¹⁷.

Заключение

Предложена лингвистическая модель календарного времени — концептуальная модель для представления семантического значения т. наз. недоопределенных темпоральных выражений, в первую очередь контекстных дат — выражений с референцией к календарному интервалу, содержащих отсылочные показатели или эллипсис. Предложен способ формального описания значения контекстных дат в виде выражений, содержащих функции. В роли переменных выступают атрибуты интервала отсчета.

Следующими шагами в рамках лингвистического анализа, необходимыми для вычисления значения контекстных дат, являются:

- разработка алгоритма автоматического семантического анализа контекстных дат на основе предложенного аппарата;
- разработка алгоритмов определения типа отсылочного показателя (анафорическая, катафорическая или дейктическая отсылка) и типа эллипсиса (анафорический или ситуативный);
- для дейктических отсылок — определение режима интерпретации (нарративный или персональный);
- установление времени отсчета.

Дальнейшие шаги входят в компетенцию формальной модели времени: определение значений атрибутов интервала отсчета и подстановка их в выражение для получения значений атрибутов интервала, соответствующего контекстной дате.

¹⁷Так, грамматическое время предиката в силу разных причин не может служить надежным ориентиром при выборе направления поиска, например: *Делегацию ждали* [прош. вр.] *в четверг* [предстоящий] (или *Делегация должна была* [прош. вр.] *приехать в четверг* [предстоящий]), *а она прибыла сегодня* [в среду].

Список литературы

- [1] Е. А. Сулейманова. «Лингвистическое моделирование темпорального адverbиала со значением локализации события», *Программные системы: теория и приложения*, **6:4(27)** (2015), с. 209–225, URL: http://psta.psir.ru/read/psta2015_4_209-225.pdf ↑^{367,377}
- [2] J. Pustejovsky, B. Ingria, R. Sauri, J. Castano, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, L. Mani, “The Specification Language TimeML”, *The Language of Time: A Reader*, eds. Mani L., Pustejovsky J., Gaizauskas R., Oxford University Press, 2005. ↑³⁶⁹
- [3] J. Strötgen, M. Gertz, “HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions”, *Proceedings of the 5th International Workshop on Semantic Evaluation*, ACL 2010 (Uppsala, Sweden, 15–16 July 2010), pp. 321–324. ↑³⁷⁰
- [4] W. Sun, A. Rumshisky, Ö. Uzuner. “Normalization of Relative and Incomplete Temporal Expressions in Clinical Narratives”, *Journal of the American Medical Informatics Association*, **22:5** (2015). ↑^{368,370}
- [5] I. Pratt, N. Francez. “Temporal prepositions and temporal generalized quantifiers”, *Linguistics and Philosophy*, **24:2** (2001), pp. 187–255. ↑³⁷⁰
- [6] R. Dale, P. Mazur, “Local Semantics in the Interpretation of Temporal Expressions”, *Proceedings of the Workshop on Annotating and Reasoning about Time and Events* (Sydney, July 2006), Association for Computational Linguistics, Sydney, 2006, pp. 9–16. ↑³⁷⁰
- [7] F. Schilder, Ch. Habel, “From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages”, *Proceedings of the ACL-2001 Workshop on Temporal and Spatial Information Processing* (Toulouse, 2001), pp. 65–72. ↑³⁷⁰
- [8] M. Negri, L. Marseglia. *Recognition and normalization of time expressions: ITC-IRST at TERN 2004*, Technical Report WP3.7, Information Society Technologies, February 2005. ↑³⁷⁰
- [9] И. В. Ефименко, «Время в мультязычных коллекциях документов: лингвистическая модель и ее реализация в среде GATE», *Девятая Всероссийская конференция по искусственному интеллекту КИИ-2004*. Т. 2, Физматгиз, М., 2004, с. 525–532. ↑³⁷⁰
- [10] Н. А. Власова. «Об одной проблеме автоматического извлечения временной информации из русскоязычных текстов», *Программные системы: теория и приложения*, **5:4(22)** (2014), с. 231–242, URL: http://psta.psir.ru/read/psta2014_4_231-242.pdf ↑³⁷⁰
- [11] А. Н. Виноградов, И. Н. Воздвиженский, Д. А. Кормалев, Е. П. Куршев. «Моделирование временного аспекта описания ситуации в задаче извлечения информации из текстов», *Программные системы: теория и приложения*, **5:4(22)** (2014), с. 215–229, URL: http://psta.psir.ru/read/psta2014_4_215-229.pdf ↑³⁷⁰

- [12] Ю. Д. Апресян. «Дейксис в лексике и грамматике и наивная модель мира», *Семiotика и информатика*, 1986, №28, с. 533. ^{↑371}
- [13] А. С. Нариньяни, «НЕ-факторы: State of Art», *Научная сессия МИФИ-2004*. Т. 3, с. 26–30. ^{↑374}
- [14] Е. В. Падучева, «Семантические явления в высказываниях от 1 лица: Говорящий и Наблюдатель», *Доклады к съезду славистов (Охрид, 2008)*, URL: http://lexicograph.ruslang.ru/TextPdf1/slavisty_2008.pdf ^{↑379}
- [15] Е. В. Падучева. *О семантике синтаксиса. Материалы к трансформационной грамматике русского языка*, Наука, М., 1974. ^{↑391,392}

Рекомендовал к публикации

к.т.н. Е. П. Куршев

Об авторе:



Елена Анатольевна Сулейманова

Научный сотрудник Исследовательского центра искусственного интеллекта ИПС им. А. К. Айламазяна РАН, одна из разработчиков технологии построения систем извлечения информации

e-mail:

yes@helen.botik.ru

Пример ссылки на эту публикацию:

Е. А. Сулейманова. «Семантический анализ контекстных дат», *Программные системы: теория и приложения*, 2015, **6**:4(27), с. 367–399.

URL: http://psta.psiras.ru/read/psta2015_4_367-399.pdf

Elena Suleymanova. *Semantic analysis of contextual dates.*

ABSTRACT. The paper suggests a notation for representing the semantics of contextual dates—cases of indirect and elliptical reference to single calendar intervals. Contextual dates are part of a large number of so-called underspecified temporal expressions. The representation is based on a linguistic model of calendar time. The metalanguage of the linguistic model is used to represent the semantic meaning of contextual dates' structural components that contain indexicals, taking into account the type of indexical, the type of component and its position in the date. Various types of ellipsis in contextual dates as well as the ways of reconstructing the omitted information are discussed. (*In Russian*).

Key Words and Phrases: natural language processing, interpretation of underspecified temporal expressions, semantic analysis.

References

- [1] E. A. Suleymanova. “Linguistic modeling of temporal adverbial localizer”, *Program systems: theory and applications*, 6:4(27) (2015), pp. 209–225 (in Russian), URL: http://psta.psiras.ru/read/psta2015_4_209-225.pdf
- [2] J. Pustejovsky, B. Ingria, R. Sauri, J. Castano, J. Littman, R. Gaizauskas, A. Setzer, G. Katz, L. Mani, “The Specification Language TimeML”, *The Language of Time: A Reader*, eds. Mani L., Pustejovsky J., Gaizauskas R., Oxford University Press, 2005.
- [3] J. Strötgen, M. Gertz, “HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions”, *Proceedings of the 5th International Workshop on Semantic Evaluation*, ACL 2010 (Uppsala, Sweden, 15–16 July 2010), pp. 321–324.
- [4] W. Sun, A. Rumshisky, Ö. Uzuner. “Normalization of Relative and Incomplete Temporal Expressions in Clinical Narratives”, *Journal of the American Medical Informatics Association*, 22:5 (2015).
- [5] I. Pratt, N. Francez. “Temporal prepositions and temporal generalized quantifiers”, *Linguistics and Philosophy*, 24:2 (2001), pp. 187–255.
- [6] R. Dale, P. Mazur, “Local Semantics in the Interpretation of Temporal Expressions”, *Proceedings of the Workshop on Annotating and Reasoning about Time and Events* (Sydney, July 2006), Association for Computational Linguistics, Sydney, 2006, pp. 9–16.
- [7] F. Schilder, Ch. Habel, “From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages”, *Proceedings of the ACL-2001 Workshop on Temporal and Spatial Information Processing* (Toulouse, 2001), pp. 65–72.
- [8] M. Negri, L. Marseglia. *Recognition and normalization of time expressions: ITC-IRST at TERN 2004*, Technical Report WP3.7, Information Society Technologies, February 2005.
- [9] I. V. Yefimenko, “Time in multilingual document collections: linguistic model and its implementation in the GATE environment”, *The 9th All-Russian Conference in Artificial Intelligence CAI-2004*. V. 2, Fizmatgiz, M., 2004, pp. 525–532 (in Russian).

- [10] N. A. Vlasova. “On one problem of automatic information extraction from Russian texts”, *Program systems: theory and applications*, 5:4(22) (2014), pp. 231–242 (in Russian), URL: http://psta.psir.ru/read/psta2014_4_231-242.pdf
- [11] A. N. Vinogradov, I. N. Vozdvizhenskiy, D. A. Kormalev, Ye. P. Kurev. “The time aspect modelling of situation description for information extraction task”, *Program systems: theory and applications*, 5:4(22) (2014), pp. 215–229 (in Russian), URL: http://psta.psir.ru/read/psta2014_4_215-229.pdf
- [12] Yu. D. Apresyan. “Deixis in lexicon and grammar, and a naive model of the world”, *Semiotika i informatika*, 1986, no.28, pp. 533 (in Russian).
- [13] A. S. Narinyani, “NOT-factors: State of Art”, *Science session MPhI-2004*, V. 3, pp. 26–30 (in Russian).
- [14] Ye. V. Paducheva, “Semantic phenomena in 1st-person utterances: the Speaker and the Observer”, *Contributions to the Congress of Slavists (Okhrid, 2008)* (in Russian), URL: http://lexicograph.ruslang.ru/TextPdf1/slavisty_2008.pdf *On the semantics of syntax. Materials for the transformational grammar of the Russian language*, Nauka, M., 1974 (in Russian)

Sample citation of this publication:

Elena Suleymanova. “Semantic analysis of contextual dates”, *Program systems: theory and applications*, 2015, 6:4(27), pp. 367–399. (In Russian).

URL: http://psta.psir.ru/read/psta2015_4_367-399.pdf