

Е. А. Сулейманова

О референциальных аспектах задачи извлечения фактов

Аннотация. В статье рассмотрены некоторые референциальные особенности новостных текстов электронных СМИ, используемых системой извлечения информации в качестве материала для фактографического поиска. Целью поиска является извлечение из текстов информации об объектах и ситуациях заданного типа — о лицах, о должностях лиц в организациях и геополитических единицах, об отношениях между лицами, а также назначениях и отставках.

Ключевые слова и фразы: извлечение информации, анализ новостей, референция, повторная номинация.

Введение

Задача поиска состоит в том, чтобы «просканировать» большой объем текста и получить в результате относительно небольшое число структур, содержащих только информацию заданного типа. Эти структуры — факты — предназначены для включения в базу фактов (часть базы знаний системы), с возможностью дальнейшего использования в аналитических целях (для вывода новых фактов, получения ответов на вопросы и т.п.).

Понимания содержания текста в целом при этом не требуется. Рассмотрим небольшой фрагмент новостного текста:

ПРИМЕР 0.1. *Федеральное агентство воздушного транспорта (Росавиация) прекратит деятельность авиакомпании ОАО «Авиалинии Дагестана», сообщил в воскресенье агентству «Прайм» глава ведомства Александр Нерадько.*

Работа выполнена при финансовой поддержке Министерства образования и науки Российской Федерации (госконтракт № 07.514.11.4109).

Предположим, в задачи системы входит поиск фактов типа «лицо-занимает-должность-в-организации». Очевидно, что приведенный в примере текст информирует о других событиях и содержит гораздо больше информации, чем можно «втиснуть» в схему факта «лицо-занимает-должность-в-организации». С другой стороны, единственный фрагмент, в котором явно говорится о лице и его должности (в примере он выделен жирным курсивом), сообщает о том, что *Александр Нерадько — глава (какого-то) ведомства*. Едва ли такой результат можно было бы считать удовлетворительным. Необходимо установить, какая организация скрывается за словом *ведомство*.

Отнесение текстовой единицы к внеязыковому объекту называется *референцией*. Сам внеязыковой объект, с которым соотносится текстовое выражение, называется его *референтом*. Для правильной интерпретации описания факта необходимо, таким образом, правильно установить референт(а) текстового упоминания участника, т.е. *разрешить референцию* текстового выражения. В тексте, кроме *ведомства*, есть еще четыре упоминания об организациях (они выделены простым курсивом). Референтом именной группы *ведомства* в нашем случае оказывается организация, которая является референтом именной группы *Федеральное агентство воздушного транспорта*, а также именной группы *Росавиация*. Все эти именные группы обладают в нашем тексте референциальным тождеством, т.е. кореферентны.

Задача установления референциального тождества (разрешения кореферентности) в ее традиционной постановке подразумевает коиндексацию кореферентных выражений. Для целей извлечения информации в нашем понимании этого недостаточно. Возникает необходимость в построении специального референциального уровня представления текста — уровня, моделирующего участников ситуаций. Единицы этого уровня — референты текстовых упоминаний, точнее модели референтов, поскольку под первыми понимаются реальные действующие лица описываемых событий. Модели референтов, получаемые в результате объединения фрагментов кореферентных описаний, и являются теми объектами, которые подлежат включению в базу фактов.

1. Референциальная связность текста в аспекте моделирования

Текст новости подчиняется общим законам построения текста и обладает свойствами *связности* и *цельности*. Цельность понимается как единая функциональная направленность. Основной целью создания текста новостного жанра можно считать сообщение фактической информации, что не исключает привнесения в текст некоторого эмоционально-оценочного компонента.

Связность текста обеспечивается как его внутренней структурой, так и внешними по отношению к тексту факторами (например, соотношением фоновых знаний автора и адресата). В теории дискурсивного анализа первый тип связности называют когезией, а второй — когерентностью [1].

Связность текста проявляется на разных уровнях. В нарративном (повествовательном) тексте выделяют референциальную (тождество участников), пространственную, временную и событийную связность [2].

Основным инструментом поддержания референциальной связности текста является *повторная номинация*. Под повторной номинацией понимают выбор средств лексической замены для уже названного субъекта или объекта [3]. К другим терминам, описывающим это явление, относятся *повторное упоминание, переименование, номинативное варьирование, контекстуальные синонимы, вторичные наименования*.

Выбор референциальных средств в тексте (референциальный выбор) определяется взаимодействием множества факторов, часто диалектически противоборствующих.

По мнению последователей когнитивных теорий референции, при построении повторной номинации автор, не забывая об экономии языковых усилий, должен сделать ее (номинацию) максимально приспособленной для того, чтобы адресат мог правильно осуществить поиск в памяти [4]. При этом механизм языковой экономии таков, что, акцентируя наиболее значимые объекты и цели, автор уделяет меньшее внимание тем, которые представляются ему менее значимыми. Основная цель автора определяется при этом как желание облегчить процесс восприятия и интерпретации сообщения адресатом. Таким образом, ведущим в коммуникации объявляется

принцип сотрудничества, определяющий ориентированность содержания и формы текста на когнитивно-психологические особенности его восприятия предполагаемым реципиентом информации [5].

Согласно коммуникативному принципу сотрудничества, или кооперации [6], референциальный выбор должен обеспечивать оптимальную возможность понимания сообщения адресатом, и только в этом случае он будет соответствовать главной цели дискурса [5]. Говоря о следовании принципам успешности коммуникации и универсальному принципу экономии языкового материала, нельзя не принимать во внимание того, что фактическое информирование — лишь одна, пусть и доминирующая, из целей новостного текста. Жанр медийных новостей несет в себе элементы публицистического функционального стиля, которому свойственны «тематическая неисчерпаемость и широта и разнообразие лексики, эмоциональность, оценочность и сила воздействия» [7]. Таким образом, стратегии экономии языковых средств и соображения легкости декодирования сталкиваются со свойственным языку вообще и художественной публицистике в частности стремлением к вариативности выражения, усилению выразительности повествования. Последний фактор особенно трудно поддается моделированию. Между тем, прагматическая установка автора на усиление выразительности или напряженности текста может подчинить себе все стороны дискурса, в том числе и типовые стратегии референциального выбора. Так, например, использование лексического повтора вместо анафорической замены может «повысить семантико-стилистическую напряженность текста и, следовательно, усилить его выразительность» [3].

Было бы неверно абсолютизировать коммуникативно-прагматические аспекты новостного текста. И в процессе создания, и в процессе понимания всякого текста участвуют два основных класса средств. Первые — это знания, закрепленные в системе языка, на разных ее уровнях, от уровня графических символов до семантики. Коммуникативно-прагматические факторы — второй класс средств — играют в этих процессах роль «практических целей, направляющих использование неязыковых знаний в сочетании с языковыми» [8].

1.1. Когнитивные модели

Большое внимание вопросам референциального выбора уделяют когнитивные подходы к анализу дискурса. Несмотря на отсутствие единого мнения о когнитивных механизмах, отвечающих за референциальный выбор, все когнитивные теории разделяют положение о том, что чем более доступен (ожидаем, активирован) референт в данной точке дискурса, «тем меньше усилий требуется для его обработки, и тем меньше формального материала затрачивается на его кодирование» (цитата из [9] приводится по [4]). Степень доступности референта в каждой точке дискурса, зависящая от комплексного действия совокупности факторов, определяет выбор референциального выражения.

Многофакторная количественная модель референциального выбора [10, 11] позволяет объяснить выбор референциального средства в каждой точке дискурса по шкале: <только полная именная группа (далее ИГ)> → <полная ИГ, местоимение> → <полная ИГ или местоимение> → <местоимение, полная ИГ> → <только местоимение>. Для этого используются эмпирически полученная референциальная стратегия и текущий интегральный коэффициент активации, вычисляемый в данной точке дискурса для каждого референта с учетом факторов активации. К факторам активации относятся линейное расстояние до antecedента, риторическое расстояние (понятие из RST — Теории риторической структуры [12]), расстояние в абзацах, семантический и синтаксический статус antecedента, одушевленность, статус протагониста (два последних фактора связаны с внутренними свойствами референта).

Когнитивные подходы несут в себе ценные идеи о влиянии дискурсивных факторов на референциальный выбор и разрешение референции, о способах количественной оценки такого влияния. При этом объяснительные и предсказательные когнитивные модели противопоставляют лексически полные и редуцированные референциальные средства. К полным относятся имена собственные и ИГ-дескрипции. Редуцированные средства — местоимения и анафорические нули — экономичны, но потенциально неоднозначны (могут иметь широкую сферу референции) и допускают возникновение референциальных конфликтов. Считается, что для предотвращения референциального конфликта в той точке дискурса, где он может возникнуть, автор текста предпочтет редуцированному средству лексически полную именную группу, что в большинстве случаев

разрешает конфликт между разными референтами-кандидатами. Таким образом, выбор референциального выражения из арсенала лексически полных средств, так же как и разрешение кореферентности дескрипций, оказывается на периферии интересов когнитивных моделей референции.

Между тем, даже беглый взгляд на тексты новостей в СМИ свидетельствует о том, что как раз лексически полные именные группы выступают в них в качестве основного средства повторной номинации. Без возможности успешного разрешения кореферентности именных групп невозможно построить по тексту новости полное описание изложенных в нем фактов (т.е. заполнить фрейм ситуации и фреймы её участников).

2. Интродуктивная и идентифицирующая референция в новостных текстах

2.1. Типы референции и референциальных средств

Для референции к тем или иным референтам говорящий (автор) может использовать два основных типа референциальных средств: лексически полные и редуцированные. Как уже говорилось, к редуцированным средствам относят анафорическое местоимение и анафорический ноль. Редуцированные средства могут использоваться для отсылки только к тем референтам, которые уже имеют к данному моменту высокую степень активации в дискурсе.

Лексически полные референциальные средства — это чрезвычайно разнородный класс языковых средств. К ним относятся имена собственные и именные группы — нарицательное имя с зависимыми или без.

Различают два типа референции: *дескриптивную* и *индексальную* [13]. Для осуществления дескриптивной референции важно понятийное содержание ИГ (дескрипции), тогда как при индексальной референции дескриптивное содержание либо вообще отсутствует (когда в качестве референциального средства используется личное местоимение), либо не является достаточным для идентификации референта.

ИГ могут использоваться для референции обоих типов и подразделяются на *дескриптивные* и *индексальные*. Дескриптивные ИГ не содержат в своем составе местоименных референциальных показателей, тип референции определяется особенностями самой

описания. ИГ, включающие местоименные референциальные показатели, относятся к индексальным.

2.2. Интродуктивная референция. Имена собственные и описания

Денотативное пространство [13] новости — это некоторый фрагмент реальной действительности, в котором должны быть фиксированы упоминаемые в тексте референты. По мере разворачивания повествования автор вводит в поле зрения адресата участников описываемых ситуаций. Одна из задач, которую при восприятии текста решает адресат, — установить референцию в каждом конкретном случае. Это означает локализовать в релевантном денотативном пространстве объект, впервые попавший в его поле зрения, или идентифицировать известный объект с некоторым объектом, хранящимся в его (адресата) памяти.

Основным средством индивидуализации и идентификации объекта, выделения его из ряда подобных, является имя собственное. Имя собственное по Расселу [14] — это простой символ, обозначающий индивидуальный предмет и обладающий этим значением вне зависимости от значений других слов.

Для того чтобы имя собственное могло выполнять в тексте идентифицирующую функцию, носитель имени должен быть известен адресату текста. Основной *прагматический принцип употребления собственных имен* заключается в том, что в отсутствие специальных показателей интродуктивности имя собственное может быть употреблено конкретно-референтно только в том случае, если адресату (по мнению автора) носитель имени собственного уже известен [13]. Известность предполагает наличие в сознании адресата «мысленного досье» носителя имени — информации в виде совокупности описаний.

Если известность референта адресату не предполагается, то, прежде чем использовать имя собственное для референции к объекту, автор должен *представить носителя имени*, т.е. дать адресату инструкцию открыть в сознании новое «мысленное досье» и предоставить некоторую информацию в форме одной или нескольких описаний для включения в это досье.

Интродуктивное употребление имени собственного предполагает, таким образом, наличие при нем некоторой описания, которая и

позволит локализовать носителя имени в денотативном пространстве.

В составе типичного для новостей интродуктивного акта имя собственное употребляется автонимно (т.е. обозначает само себя) в синтаксической функции приложения при дескрипции. Для названий организаций характерна позиция несогласованного приложения (*деятельность авиакомпании ОАО «Авиалинии Дагестана», крупнейший ритейлер Walmart*); для геополитических единиц — несогласованного или согласованного приложения (*администрация города Реутов, мэр города Челябинска*).

Наряду с этой компактной схемой интродукции встречаются более развернутые — аппозитивные конструкции с обособленным приложением:

ПРИМЕР 2.1. [...] *Сергея Анатольевича Жвачкина — генерального директора общества с ограниченной ответственностью «Газпром Трансгаз Краснодар», Сергея Евгеньевича Ильиных — председателя Думы города Томска, Петра Савельевича Чубика — ректора федерального образовательного учреждения высшего профессионального образования «Национальный исследовательский Томский политехнический университет», [...]*

ПРИМЕР 2.2. *Крупнейший отечественный авиаперевозчик — компания «Аэрофлот» — по итогам этого года демонстрирует хорошие результаты.*

Возможна интродукция через высказывание идентификации или высказывание с предикатной ИГ, когда первый компонент — имя собственное, а второй — дескрипция, которую адресат текста должен «подшить» в «мысленное досье», открываемое в этот момент на носителя имени:

ПРИМЕР 2.3. *ОАО ЛОРП — крупнейшее предприятие водного транспорта северо-восточного региона страны.*

Иногда всё же допускаются отступления от нейтральной схемы представления действующих лиц.

Имя собственное может быть введено в текст **без представления**, если его носитель предполагается известным читателю (т.е. основной прагматический принцип при этом соблюдается). Без интродуктивной дескрипции (или с «тикетным» именем нарицательным, которое не относится к интродуктивным дескрипциям в

силу семантической неполноты) могут быть впервые упомянуты общеизвестные геополитические единицы и организации, политические и государственные деятели, а также «медиа-лица» и прочие персоны, чьи имена предполагаются известными читателю (т.е. носители так называемых «прецедентных имен»):

ПРИМЕР 2.4. *«Единая Россия» определилась со списком кандидатов на пост губернатора Московской области.*

ПРИМЕР 2.5. *Президент Путин подписал указ [...]* (этикетное употребление)

Довольно редко в новостях встречается *имплицитное представление*, при котором первому упоминанию имени собственного предшествует дескрипция. В таком случае автор предполагает, что читатель установит кореферентность дескрипции и имени собственного на основании презумпции связности текста:

ПРИМЕР 2.6. *Адвокат Михаила Ходорковского надеется, что [...]. Об этом Юрий Шмидт заявил в интервью радиостанции "Коммерсант FM".*

Одной из особенностей рассматриваемой в задаче предметной области является наличие довольно большого числа сущностей, которые вообще не именуется с помощью имен собственных. К объектам такого рода относятся все должности и некоторые типы организаций (*администрация, правление, совет директоров* и т.п.). Такие сущности индивидуализируются через отношения с другими сущностями. На языковом уровне это проявляется в наличии у лексемы сильной валентности, например: *директор, глава, руководитель (чего?), заместитель, помощник, советник (чей?), администрация (чего?), администрация (чья?)*.

Кроме того, для индивидуализации таких сущностей могут служить дополнительные условия — заполненная «ограничительная» валентность (*директор по персоналу, заместитель по научной работе, министр здравоохранения и социального развития*) или признак (*первый вице-президент, коммерческий директор*).

Для группы целевых объектов — организаций характерна неразличимость дескрипции и имени собственного, например: *Федеральная налоговая служба, Министерство Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий, Запорожский алюминиевый*

комбинат. Как правило, у таких объектов имеются сокращенные наименования, которые функционируют как имена собственные (*ФНС России, МЧС России, ЗАЛК*).

Коммерческим организациям свойственно включать дескрипцию в состав полного (официально зарегистрированного) названия, например: *ОАО «Запорожский алюминиевый комбинат», ОАО «Завод по производству систем программного управления»*. Имя собственное в узком смысле тоже может быть встроено в такое официальное название, что часто приводит к появлению непарных кавычек: *Открытое акционерное общество «Нефтегазовая компания «Славнефть» (ОАО «НГК «Славнефть»)*.

Геополитические единицы некоторых типов — области, края, республики, федеральные округа и др. — индивидуализируются не именами собственными, а добавлением признака-идентификатора к общему имени (*Ярославская область, Краснодарский край, Чеченская республика, Центральный федеральный округ*).

3. Идентифицирующая референция. Повторные номинации

Из всего арсенала средств для последующей референции к введенным в рассмотрение объектам в новостных текстах заметно преобладают имена собственные и дескриптивные ИГ.

Изменения, которым подвергаются исходные номинации при последующих употреблениях в идентифицирующей функции, можно разделить на два вида: *количественная редукция и замена*.

Кроме того, будем различать характер выбора повторной номинации объекта — *эксплицитный* или *имплицитный*.

3.1. Эксплицитный и имплицитный выбор номинации

Эксплицитный выбор имеет место, если для получения повторной номинации автор не прибегает ни к каким знаниям о самом объекте, кроме явно указанных в тексте и тех, которые могут быть получены из них на основе лингвистических и онтологических знаний (общих знаний об устройстве мира и предметной области). Если же номинация ссылается на знания об объекте из «мысленного досье» автора, не упомянутые в тексте, то такой выбор будем называть имплицитным.

Имплицитный выбор имеет место при так называемой «презумптивной» определенной референции [13]. Употребляя некоторую

дескрипцию носителя имени собственного из своего досье, автор предполагает, что такая же дескрипция имеется в досье адресата (как в случае с «прецедентными именами»), например:

ПРИМЕР 3.1. *Адвокат Михаила Ходорковского надеется, что освобождение по амнистии большого количества заключенных, осужденных за экономические преступления, и в том числе экс-главы ЮКОСа, возможно по специальному постановлению Госдумы РФ.*

Однако автор может ввести в повторную номинацию заведомо новую для адресата информацию, предполагая, что последний установит кореферентность, основываясь на презумпции связности текста.

Поскольку не существует жесткой границы между онтологией (концептуальными знаниями) и базой фактов (знаниями об экземплярах), то различие между имплицитными и эксплицитными номинациями тоже не абсолютно (например, не очевидно, имплицитным или эксплицитным считать выбор дескрипции *столица* для экземпляра *Москва* или дескрипции *президент США* для лица по имени *Барак Обама*).

3.2. Количественная редукция и замена

Будем считать это противопоставление комплементарным: замену определим как такое изменение исходной номинации, которое не является количественной редукцией.

Количественной редукцией будем называть любые формы сокращения состава номинации путем удаления тех или иных элементов.

Имена собственные, состоящие из нескольких слов, в функции повторной номинации могут подвергаться различным видам редукции. Личные имена собственные могут выступать в более или менее полной форме, например: [*<имя—отчество—фамилия>* или *<имя—среднее имя (средние имена)—фамилия>*] → *<имя—фамилия>* → *<фамилия>*. Для сокращения многословных названий или названий-дескрипций используются следующие виды редукции и их комбинации: замена инициальной аббревиатурой (*Воронежское акционерное самолетостроительное общество — ВАСО*), акронимом (*Минфин*), сложносокращенным названием (*Государственная дума*

— *Госдума, Законодательное собрание — Заксобрание*) или простым сокращением без усечения (*МЧС России — МЧС*).

Замене могут подвергаться как имена собственные, так и дескрипции.

Имена собственные допускают синонимические замены — это свойственно геополитическим единицам (*Республика Саха — Якутия*).

Для повторной референции к геополитическим единицам и организациям характерно использование общих имен вместо названий-дескрипций. При этом, поскольку собственное имя у таких объектов часто представляет собой соединение общего имени и признака-идентификатора, внешне такая замена может выглядеть как чисто количественная редукция: *Ярославская область — область, Краснодарский край — край, Чеченская республика — республика, Центральный федеральный округ — округ, Республиканская партия — партия, Министерство обороны — министерство, Федеральная служба по надзору в сфере образования и науки — служба*. Характерная позиция для таких номинаций — позиция генитивного зависимого в составе другой дескрипции (*глава службы, руководитель министерства, губернатор края*). Индексальная референция при повторном упоминании таких референтов практически не используется.

Наибольшим разнообразием в текстах отличаются номинации лица. Это отражает тот факт, что человек значительно многограннее в своих проявлениях, чем прочие объекты, и в языке получает «множество разных обозначений. Он может быть назван по своим общественным функциям, взглядам, моральному облику, нравственным склонностям и вкусам, по своим поступкам и поведению, семейному статусу, родственным связям, внешнему виду, участию в тех или иных событиях и происшествиях, отношению к нему говорящего и многому другому» [15]. Наиболее распространены в качестве номинаций лица функциональные имена (названия лиц по должности, роду занятий, титулы, ранги, звания), реляционные имена (например, термины родства) и актуальные имена (носители ситуативного признака, например *кандидат* в ситуации выборов).

Перечислим основные виды замены, характерные для повторных номинаций в новостных текстах:

- (1) замена ИГ на ИГ, которой соответствует вышестоящий концепт: *губернатор — глава, область — регион, дума — парламент*;

- (2) замена имени экземпляра (имени собственного или названия-описания) на ИГ, которой соответствует родительский концепт экземпляра: *МЧС — министерство, Приморский край — край*;
- (3) замена ИГ с семантикой базового концепта на ИГ с семантикой аспекта: *компания — ритейлер*;
- (4) замена-трансформация: *администрация края — краевая администрация, губернатор Приморья — приморский губернатор*;
- (5) «метонимическая» замена: *адвокат Х-а* (реляционное имя) — *адвокат* (род занятий); вариант такой замены — подмена референта, например: *Московская область — Подмоскowie*;
- (6) синонимическая и квазисинонимическая замена: *адвокат* (род занятий) — *правозащитник, мэр — градоначальник, городской голова*;
- (7) ассоциативная замена: *премьер-министр — политик, министр — чиновник, федеральное агентство — ведомство*.

Эти виды замены могут комбинироваться друг с другом: *мэр Москвы — столичный градоначальник* (промежуточное звено: *Москва — столица*).

Иногда встречается прономинализация:

ПРИМЕР 3.2. *На Ставрополье назначен председатель правительства края. Им стал Юрий Трутнев.*

Довольно распространено явление, когда в денотативном пространстве новости (на которое указывает время её публикации) объект уже не обладает «описательным» свойством. В таком случае в описание включается специальный лексический показатель: *бывший, экс-, прежний*.

Заметим, что это довольно нетривиальное явление с точки зрения представления в результатах извлечения.

4. Заключение

Проблемы референции и установления референциального тождества относятся к числу наиболее сложно поддающихся моделированию. Для адекватного описания этих явлений «требуется решение многих теоретических вопросов, каждый из которых представляет самостоятельную сложную проблему. Это, прежде всего, проблема соотношения лингвистического и экстралингвистического в процессе

именования и переименования объектов и явлений, проблемы соотношения номинации и референции, значения и обозначения, формы и функции» [16].

Тем не менее, несмотря на теоретическую сложность и обширность проблемы, мы надеемся достичь практически приемлемых результатов в алгоритмическом разрешении индексальной и дескриптивной референции для довольно узкого класса задач фактографического поиска. Необходимым шагом на этом пути нам представляется описание характерных для текстов предметной области способов и схем повторной номинации к целевым объектам.

Список литературы

- [1] Кронгауз М. А. Семантика. М : РГГУ, 2001. — 299 с. ↑1
- [2] Givón T. *Coherence in text, coherence in mind* // Pragmatics and cognition, 1993. Vol. 1(2) ↑1
- [3] Валгина Н. С. Теория текста: Учебное пособие. М. : Логос, 2003. ↑1
- [4] Фёдорова О. С. *Текстовая анафора: сочетание статистического и когнитивного подходов (на материале цахурского языка)* // Труды Международного семинара Диалог-2000 по компьютерной лингвистике и ее приложениям. — Протвино, 2000. Т. 1 ↑1, 1.1
- [5] Сергеева М. Э. *Принцип языковой экономии и его отражение на уровне орфографии (на материале английского языка)*, http://www.bigpi.biysk.ru/in yaz/readarticle.php?article_id=7 ↑1
- [6] Грайс Г. *Логика и речевое общение*. Новое в зарубежной лингвистике. Вып. XVI—М. : Прогресс, 1985 ↑1
- [7] Кибрик А. А. *Обосновано ли понятие "Дискурс СМИ"?* // Жанры и типы текста в научном и медийном дискурсе. — Орел : ОГИИК, 2008, с. 6-11 ↑1
- [8] Городецкий Б. Ю. *Компьютерная лингвистика: моделирование языкового общения* : Вступ. ст. Новое в зарубежной лингвистике. Вып. XXIV—М. : Прогресс, 1989 ↑1
- [9] Givón T. *Topic continuity in discourse: A quantitative cross-language study* // The Philosophy of Language / ed. T. Givón— Amsterdam : Benjamins, 1983 ↑1.1
- [10] Кибрик А. А. *Референция и рабочая память: о взаимодействии лингвистики с психологией и когнитивной наукой* // Материалы первой российской интернет-конференции по когнитивной науке. — М. : УМК Психология, 2004, с. 29-43 ↑1.1
- [11] Кибрик А. А. *Анализ дискурса в когнитивной перспективе*, Дис... докт. филол. наук, Институт языкознания РАН, 2003. — 90 с. ↑1.1
- [12] Mann W.C., Thompson S.A. *Rhetorical Structure Theory: A Theory of Text Organization* // Technical Reports ISI/RS-87-190, 1987 ↑1.1
- [13] Шмелёв А. Д. *Русский язык и внеязыковая действительность*. М. : Языки славянской культуры, 2002. — 496 с. ↑2.1, 2.2, 3.1

- [14] Russel B. *Descriptions* // *The Philosophy of Language* / ed. A. P. Martinich — Oxford : Oxford University Press, 1919 ↑2.2
- [15] Арутюнова Н. Д. Предложение и его смысл: Логико-семантические проблемы. М. : Наука, 1976. ↑3.2
- [16] Фатеева Е. Д. *Переименование в лингвистике текста*, Автореф. дис... канд. филол. наук, МГПИ им. В.И.Ленина, 1983. — 16 с. ↑4

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

Образец ссылки на эту публикацию:

Е. А. Сулейманова. *О референциальных аспектах задачи извлечения фактов* // Программные системы: теория и приложения : электрон. научн. журн. 2012. Т. 3, № 3(12), с. 41–56.

URL: http://psta.psiras.ru/read/psta2012_3_41-56.pdf

E. A. E. A. Suleymanova. *On referential aspects of fact extraction.*

ABSTRACT. The article discusses some reference-related features of Web media news articles, viewed as the data source for an information extraction system. The target information includes persons, their positions in organizations and geopolitical entities, appointments and dismissals, interpersonal relationships. (*in Russian*).

Key Words and Phrases: information extraction, news texts, reference, referent mentions.