



С. Р. Момот, И. Н. Воздвиженский

Оценка фактуальности для пропозиции в синтаксически подчиненном положении

Аннотация. В статье описывается методика определения фактуальности пропозиции для целей автоматического анализа текстов. Особое внимание уделено пропозиции, которая находится в синтаксически подчиненном положении. Дается определение понятию «фактуальный статус» и описываются его компоненты, основным из которых является т.н. «эпистемическая оценка».

Методика заключается в построении иерархии предикативных вершин в полипредикативном предложении. На основании этой иерархии для каждой выделенной предикативной вершины с помощью правил и словаря устанавливается эпистемическая оценка.

Ключевые слова и фразы: фактуальность, фактуальный статус, эпистемическая оценка, временной план, гипотаксический контекст, предикат, пропозиция.

Введение

Оценка фактуальности событийной информации, то есть способа ее соотношения с реальностью, актуальна для различных прикладных задач, связанных с анализом событийной информации в тексте. Надежное решение в этой области может быть полезно и для целей непосредственного извлечения событийной информации, так как позволит отделить упоминания о событиях, не соотносимых с реальностью, от событийной информации, поданной как реальная или возможная.

Предлагаемое решение основано на анализе поверхностно-синтаксических связей с помощью правил и словаря. Объектом анализа являются пропозиции в пределах одного предложения — точнее говоря, входящие в их состав семантические предикаты как центральный компонент пропозиции. Целевой пропозицией, подлежащей оценке, является событийная пропозиция в предложении. Событийная пропозиция портретирует происходящее в действительности, тогда как

Работа выполнена в рамках НИР «Исследование и разработка методов интеллектуального анализа данных», номер гос. регистрации АААА-А19-119020690042-2.

© С. Р. Момот, И. Н. Воздвиженский, 2020

© Институт программных систем имени А. К. Айламазяна РАН, 2020

© Программные системы: теория и приложения (дизайн), 2020



противоположностью событийной выступает логическая пропозиция, которая сообщает о результатах неких умственных операций или об установлении определенных признаков.

Для того чтобы использовать событийную информацию с размеченной фактуальностью в прикладных целях, необходимо определить, какие именно грамматические значения следует охватить понятием «фактуальность», какие значения фактуальности присваивать единицам текста и разработать метод для разметки этой информации в текстах.

Для определения понятия «фактуальность» и ее значений создаются системы разметки и аннотированные корпуса, преимущественно на английском языке. Самыми известными являются Factbank [1] и MEANTIME [2]. В корпусе Factbank каждому событию (размеченного по принципу языка разметки TimeML) присваивается одно из следующих значений фактуальности: CT (certain), PR (probable), PS (possible), Uu (unknown), а также полярность (+ или -). Похожие значения фактуальности используются в MEANTIME: Fact, Counterfact, Possibility (uncertain), Possibility (future). Некоторые системы разметки опираются на недискретную шкалу, например система разметки, разработанная в Вашингтонском университете. Согласно этой разметке, каждому событию присваивается значение в пределах шкалы [-3,0, 3,0], где значение «-3,0» означает «событие не имело места», а «3,0» — «событие имело место».

Автоматическое прогнозирование значения фактуальности в соответствии с выбранной шкалой может быть реализовано либо на основании машинного обучения с опорой на поверхностные признаки ([2], [3] и другие), либо на основании правил, анализирующих более глубокие грамматические признаки. Сопоставить результативность этих методов затруднительно, так как они опираются на разные системы разметки — что в свою очередь означает разную интерпретацию взаимосвязанных понятий (таких как «фактуальность», «отрицание», «эпистемическое обязательство» и др). Самые популярные методы автоматического прогнозирования фактуальности, разработанные для английского языка, сопоставлены в [4]. Для этого авторами статьи разные шкалы разметки фактуальности приведены к единой. Как показало сравнение, оптимальный результат показали два метода, основанные на правилах и использующих систему разметки Factbank: De Facto [5] и TruthTeller [6]. Оба алгоритма опираются на заранее составленные словари предикатов, которые способны подчинять пропозиционный актант (в DeFacto они называются ESP, event selecting predicate) и анализируют синтаксическое дерево предложения по направлению «сверху —

вниз», начиная с верхнего предиката к зависимому от него предикату и т.д. Отличительной особенностью алгоритма De Facto является особое внимание к источникам, цитируемым в тексте. Поскольку источники могут по-разному представлять фактуальность того или иного события, фактуальность события определяется всегда в связи с конкретным источником. Что касается алгоритма TruthTeller, то итоговая фактуальность каждого предиката определяется как функция от двух показателей: Clause Truth (фактуальность предикативной группы, определяемая по вышестоящему предикату) и Predicate Truth (фактуальность предиката, определяемая показателями модальности и отрицания внутри предикативной группы).

Подход, предлагаемый в настоящей статье, отличается от описанных выше тем, что:

- обладает собственным, ориентированным на конкретную задачу алгоритмом для построения иерархии предикатов, так как доступный синтаксический анализатор не предоставляет непосредственной связи между предикатами;
- использует словарь предикатов и наборы признаков для русского языка;
- выделяет три отдельных компонента фактуальности: эпистемическая оценка, временной план и отношение автора к источнику;
- трактует «полярность» как отдельную характеристику пропозиции, не смешивая ее с «контрфактуальностью».

1. Теоретическая база и методика исследований

1.1. Пропозиции в синтаксически независимых и синтаксически подчиненных позициях

Для того чтобы решить вопрос об отношении пропозиции к реальности, следует различать пропозиции, находящиеся в синтаксически независимом и синтаксически подчиненном положении.

Пропозиция в синтаксически независимом положении может быть находится в ассертивном (повествовательном), побудительном или вопросительном предложении. В самом базовом случае пропозиция в синтаксически независимом положении в повествовательном предложении выражена глаголом в индикативе, например: «У самолета в небе над Лос-Анджелесом *загорелся* двигатель» (предикат пропозиции выделен). Задача установления соотношения с реальностью для такой пропозиции относительно проста. Побудительные и вопросительные предложения в данной статье не рассматриваются.

Однако, как замечает Е. В. Падучева [7], пропозиция в синтаксически подчиненном положении в некоторых случаях может быть соотнесена с реальностью. Чтобы это было возможно, на фактуальность пропозиции должен указывать подчиняющий эту пропозицию предикат. Опираясь на подчиняющий предикат, можно установить фактуальность пропозиции, выраженной не только личной формой глагола (который обладает категорией наклонения, отражающей его связь с реальностью), но и неличными формами глагола (инфинитив, деепричастия) или отглагольными существительными [8].

Таким образом, для правильной оценки фактуальности события, упоминание о котором находится в синтаксически подчиненном положении, следует проанализировать не только целевую пропозицию, но и фактуальность других пропозиций, связанных с ней.

Поскольку речь идет об оценке языковых выражений в синтаксически подчиненном положении, в статье рассматриваются полипредикативные предложения. Отталкиваясь от конкретной языковой репрезентации пропозиций, с помощью правил осуществляется выделение семантических предикатов присутствующих в предложении пропозиций и устанавливаются связи между ними. Впоследствии эта картина взаимосвязанных семантических предикатов позволит оценить их взаимовлияние в ракурсе фактуальности и дать оценку целевому предикату.

1.2. Фактуальный статус пропозиции

Дать характеристику фактуальному статусу пропозиции значит дать оценку широкому спектру модальных значений.

В первую очередь необходимо определить модальность пропозиции по признаку «реальность/нереальность». Традиционно принято считать, что противопоставление по этому признаку характеризует объективную модальность [9], однако для адекватной оценки фактуальности необходимо также учитывать и некоторые субъективные элементы высказывания. Субъективный компонент, который присутствует в любом высказывании — это наличие самого говорящего (или автора текста), который берет на себя полное или неполное эпистемическое обязательство. Другими словами, в семантике высказывания присутствует имплицитный компонент «Я знаю, что...». Если говорящий берет на себя неполное эпистемическое обязательство (то есть характеризует сообщаемую информацию как неточную или сомнительную), то в высказывании появляются эксплицитные показатели субъективной модальности.

Кроме того, в тексте могут присутствовать указания на источники сообщаемой информации или путь доступа автора текста к сообщаемой информации (категория эвиденциальности), причем автор текста может указывать нам на степень надежности этих источников и высказывать доверие или недоверие по отношению к этим источникам. В том случае, если автор указывает на путь доступа к сообщаемой информации, он частично снимает с себя эпистемическую ответственность. Как отмечает Е. В. Падучева [10], прямая эвиденциальность, то есть сообщение о событии, которое говорящий сам видел или принимал в нем участие, противопоставлена косвенной эвиденциальности. В случае косвенной эвиденциальности говорящий либо основывается на данных, которые ему сообщили, либо сообщает результат своих умозаключений, либо сообщает то, что ему показалось.

Как видно, ни понятие «объективная модальность», ни понятия «эпистемическая» или «эвиденциальная модальность» недостаточны для полноценной характеристики событийной информации с точки зрения ее соответствия реальности. Это ставит вопрос о новом понятии, которое объединяло бы все эти элементы. Де-факто таким понятием стала «фактуальность», однако общепринятого определения для него не существует. В сфере автоматической обработки языка получила распространение формулировка, предложенная Р. Саури и Дж. Пустейовски: фактуальность это уровень ответственности «релевантных источников по отношению к фактуальной природе событий, упоминаемых в дискурсе» [1]. Стоит упомянуть также понятие «пропозициональная модальность», которое Палмер определяет как «отношение говорящего к фактуальному статусу пропозиции» и которое охватывает эпистемическую и эвиденциальную модальности [11].

Итак, для того чтобы определить фактуальный статус пропозиции в его полноте, следует учитывать следующие взаимосвязанные факторы:

Объективно-модальные: в основном это подразумевает оппозицию индикатива и сослагательного наклонения. Эта оппозиция традиционно относится к сфере объективной модальности, однако всегда следует принимать во внимание, что в каждом высказывании имплицитно присутствует субъект речи (автор текста или говорящий). Поэтому под «объективностью» понимается не некое истинное положение вещей, а лишь его «субъективный аналог — ощущение субъекта, что пропозиция Р является истинной» [12].

Субъективно-модальные: эксплицитно выраженная мера эпистемической ответственности автора текста или цитируемого им источника (эпистемически ответственного субъекта).

Эвиденциальные: имеют место, если автор текста, являющийся основным субъектом речи, упоминает другой субъект и сообщает о чём-либо со ссылкой на этот субъект. Необходимо установить:

- (а) дистанцируется ли автор от эпистемической оценки, которую взял на себя цитируемый им субъект или разделяет ее (это значение маркируется частицами «якобы» и т.п., а также предикатами, которые «перекладывают» эпистемическую оценку на цитируемый субъект: «*Кудрин считает, что американская экономика ослабнет*».);
- (б) выражает ли автор недоверие к самому источнику («по слухам», «согласно непроверенной информации»).

Смежные явления, не входящие в понятие «фактуальность», однако необходимые для адекватной оценки фактуального статуса пропозиции:

- (а) полярность (входит в пропозицию) — подразумевается полярность глагольных форм, выраженная частицей «не»;
- (б) временной план — противопоставляет прошлое и настоящее с одной стороны будущему с другой стороны. Противопоставление будущего времени прочим временным значениям объясняется тем, что пропозиции, выражающие события, которые еще не наступили, не отражают реальность, тем самым приближаясь к модализованным значениям. По замечанию Н. Д. Арутюновой, «будущее постоянно взаимодействует с модальными категориями» [13]. Следует подчеркнуть, что временной план определяется либо относительно момента речи, либо относительно грамматической точки отсчета, в результате чего разные пропозиции, осуществляющие референцию к одному и тому же событию в реальности, могут иметь разный временной план в пределах одного текста. В рассматриваемой модели временной план всегда оценивается в пределах одного предложения.

Соответственно, полностью охарактеризовать фактуальный статус пропозиции можно следующим образом. Для каждой событийной пропозиции устанавливается источник, который вводит ее в текст, причем для событийной пропозиции фиксируется ее полярность (положительная или отрицательная). Собственно фактуальный статус характеризуется через три компонента:

Эпистемическая оценка (ЭО). Этот компонент отражает ту меру, в которой говорящий или цитируемый субъект берет на себя эпистемическую ответственность за соответствие сообщаемого действительности. В данном случае термин «эпистемическая оценка» следует

понимать широко, так как он объединяет и значения, традиционно понимаемые как объективно-модальные (по причинам, описанным выше), и эксплицитно выраженную оценку вероятности наступления события, и неполную осведомленность автора или цитируемого субъекта о сообщаемом.

Временной план: будущее или небудущее.

Отношение автора к источнику: наличие или отсутствие дистанцирования автором от сообщаемого или от источника.

1.3. Эпистемическая оценка

Из трех перечисленных компонентов (эпистемическая оценка, временной план, отношение автора к источнику) ключевым является первый. Следует более подробно описать значения, которые он может принимать и порядок их определения.

В первую очередь следует отделить все пропозиции с «нулевой фактуальностью», т.е. никак не соотносимые с реальностью (например, нереализованные возможности в прошлом, распоряжения в повелительном наклонении, условия и т.п.).

Пропозиции с ненулевой фактуальностью можно разделить на фактуальные, контрфактуальные и нефактуальные (см. рисунок 1).

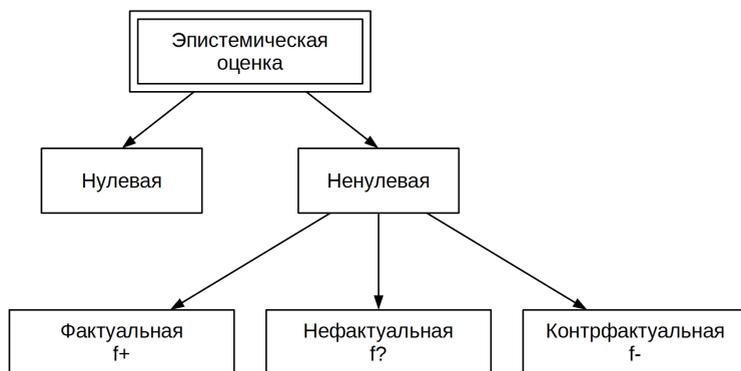


РИСУНОК 1. Возможные значения эпистемической оценки

Фактуальная ЭО ($f+$) означает, что эпистемически ответственный субъект (говорящий или цитируемый им субъект) берет на себя полную ответственность в том, что сообщаемое соответствует действительности.

Контрфактуальная ЭО (f-) означает, что пропозиция находится в сфере действия контекста, который предполагает ее несоответствие действительности.

Следует сделать два замечания касательно значения «f-». Во-первых, мы различаем контрфактуальную ЭО и отрицательную полярность предиката. Отрицательная полярность это характеристика предиката, тогда как контрфактуальная ЭО характеризует подчиненную пропозицию в целом или некоторые ее элементы (необязательно предикат). Во-вторых, в данном случае понятие «контрфактуальность» не сводится к контрфактуальным условиям («Если бы ты позвонил вчера. . .»), в отличие от традиционного использования этого термина в зарубежной лингвистике. Контрфактуальные контексты могут быть различны по своему значению: они могут предполагать отсутствие эпистемической ответственности за истинность подчиненной пропозиции, отказ субъекта совершать действие, названное подчиненной пропозицией и т.д. Кроме того, контексты различаются по сфере влияния: могут распространяться на всю пропозицию или ее часть и т.п. Однако этими различиями можно пренебречь, для целей извлечения информации важнее всего то, что их объединяет: контрфактуальность подчиненного контекста.

Нефактуальная ЭО (f?) подразумевает частичную эпистемическую ответственность, причем степень этой ответственности может быть ранжирована (например, на числовой шкале от 0 до 1).

Основным значением, относительно которого определяются остальные, является фактуальная ЭО. В каноническом случае пропозиция с фактуальной ЭО реализована в предложении полной предикативной конструкцией в автономном (не гипотаксическом) контексте, вершина которой является личной формой глагола, его временной план «небудущее». Пример канонической фактуальной ЭО (предикативная вершина пропозиции выделена): «У самолета в небе над Лос-Анджелесом *загорелся* двигатель».

Представляет интерес задача определения фактуального статуса для пропозиций в синтаксически подчиненном положении.

Предлагаемый способ анализа пропозиции с целью установления его эпистемической оценки требует перехода от поверхностно-синтаксического уровня к глубинно-синтаксическому. Конечно, требуется смоделировать лишь часть глубинной структуры, которая позволит определить ЭО целевой пропозиции. Для определения ЭО следует выделить входящие в предложение предикативные, полупредикативные и

свернуто-предикативные конструкции и установить, каковы отношения между ними. Таким образом, выражение «ЭО предиката» подразумевает ЭО предикативной вершины некой предикативной, полупредикативной конструкции или носителя свернутой предикативности.

Отдельно следует сказать о конструкциях с предикатным актантом (КПА), представляющих интерес для частичного описания глубинной структуры. Это конструкции, актантом которых непосредственно является целевая пропозиция или другая подобная конструкция (в конечном счете эта цепь подчинений приводит к целевой пропозиции). Для оценки ЭО целевой пропозиции интерес представляет определенный круг предикатов, которые, как правило, являются вершиной КПА (другими словами, в предложении выступают как включающие предикаты, а предикативная вершина их актанта в этом случае — включенный предикат). Эти предикаты перечислены в «Словаре подчиняющих предикатов», который составлен вручную и содержит около 200 записей — преимущественно глаголов, но среди них есть также отглагольные существительные и некоторые другие выражения.

Можно выделить такие группы предикатов по их семантике (список неполный, приведены только самые большие группы):

- предикаты цитирования и речи (*«сказать»*, *«сообщить»*...)
- предикаты восприятия (*«увидеть»*, *«заметить»*, *«обнаружить»*...)
- предикаты ментального действия (*«вспомнить»*, *«забыть»*, *«понять»*...)
- фазовые предикаты (*«продолжаться»*, *«длиться»*, *«ускоряться»*, *«завершиться»*), причем предикаты со значением «начала» формируют отдельную подгруппу (*«начаться»*, *«открыться»*...)
- предикаты эмоциональной реакции (*«поблагодарить»*, *«рад»*, *«удивлен»*...)
- предикаты с каузативным значением (*«заставить»*, *«спровоцировать»*...)

Для каждого предиката в словаре указаны ограничения, при условии выполнения которых этот предикат проецирует заданное значение ЭО на подчиненную пропозицию.

Эти ограничения могут заключаться в:

- семантике самого предиката (например, ряд предикатов, содержащих в семантике указание на контрфактуальность подчиненной пропозиции, такие как *«отменить»*, *«опровергнуть»*, *«предотвратить»*);

- синтаксических и морфологических характеристиках предиката (полярность, временной план, для причастий — краткая форма и т.п.)
- типе связи, которым предикат связан с подчиненной пропозицией (подробнее об этом см. ниже).

Проецируемые значения ЭО могут быть следующие:

- Фактуальная ЭО подчиненного предиката
- Контрфактуальная ЭО подчиненного предиката
- Нефактуальная ЭО подчиненного предиката
- Абсолютно-фактуальная ЭО
- Абсолютно-контрфактуальная ЭО
- ЭО неопределима

Проецируемые значения «абсолютно-фактуальная ЭО» и «абсолютно-контрфактуальная ЭО» характеризуются тем, что это значение ЭО включенного предиката остается соответственно «фактуально» или «нефактуально» даже тогда, когда собственная ЭО включающего предиката отлична от фактуальной (например, если этот предикат сам входит в сферу влияния другого включающего предиката). Значения «абсолютно-фактуальная» и «абсолютно-контрфактуальная» проецируются фактивными предикатами. Фактивные предикаты предполагают, что говорящий (или другой эпистемически ответственный субъект) гарантирует истинность пропозиции, подчиненной такому предикату (фактивными являются, к примеру, *«извиниться»*, *«изумить»*, *«поблагодарить»* и многие другие) [12].

1.4. Язык правил для анализа синтаксических отношений

Непосредственный анализ текста осуществляется с помощью правил, основная цель которых заключается в том, чтобы построить модель глубинной синтаксической структуры предложения и на основании этой структуры установить фактуальный статус каждого предиката. На вход правила получают иерархическую синтаксическую модель предложения, полученную в результате обработки предложения синтаксическим анализатором MaltParser¹. Эта синтаксическая модель представляет собой сочетание древовидной структуры, списка слов исходного текста и множества атрибутов на каждом слове исходного текста и каждом узле дерева.

Определим ряд понятий, используемых далее:

¹<http://www.maltparser.org/>

Фрагмент — непрерывный фрагмент исходного текста, который однозначно идентифицируется позицией первого и последнего своего символа в тексте. Границы фрагментов не пересекаются.

Узел — вершина дерева, описывающего зависимости между фрагментами. Каждый узел ссылается строго на один фрагмент текста. Узел может быть независимым (корневой узел) или быть подчинен строго одному вышестоящему (родительскому) узлу. Узел может иметь произвольное число непосредственных подчиненных узлов (потомки), по отношению к каждому из которых он будет являться родительским.

Родительский узел или «родитель» — ближайший вышестоящий узел в иерархии относительно данного узла.

Корневой узел или *корень* — узел, у которого нет родительского узла.

Дерево — множество узлов, восходящее по иерархии к одному единственному корневому узлу.

Атрибут — ассоциативная пара «название — значение», описывающая какой-либо один признак узла. В атрибутах преимущественно содержится морфологическая информация (часть речи, лицо, число и т.п.) Название имеет вид непрерывной (без пробелов) строки, значение может иметь различные типы (строка, число, логическое значение, узел дерева и т. п.).

Набор атрибутов — множество атрибутов, где каждый атрибут в рамках этого множества имеет уникальное название.

Контейнер узлов — хранилище узлов, отвечающее за хранение узлов в памяти и их нумерацию.

Пространство узлов контекста — это контейнер узлов, независимый от узлов исходного дерева и предназначенный для создания новых произвольных структур в ходе анализа текста.

Правила написаны на созданном для этой цели языке правил для анализа синтаксических отношений (сокращенно ЯПСО или SPSL). Одно правило SPSL описывает условие, которое должно выполняться на определенном узле дерева синтаксической структуры, связанных с ним узлов и соответствующих им фрагментов, для того, чтобы отработал набор команд, предназначенных для этого случая. Соответственно выделяются проверочная часть правила, формулирующая условие, и командная часть правила, описывающая действия при выполнении условия.

Большинство правил, написанных для данного исследования, в проверочной части обращаются к таким атрибутам узлов как часть

речи («POS»), время («Tense», для глаголов) и синтаксическая роль («parser_role») и некоторым другим. Последний из упомянутых атрибутов содержит тип синтаксического отношения, связывающего узел, содержащий этот атрибут, с его непосредственным родителем². Рассматривается одновременно текущий узел, его родитель, родитель этого родителя и т.д. до корня дерева. В командной части узлам присваиваются новые атрибуты (к которым можно обратиться в дальнейшем), избранные узлы сохраняются в пространство узлов контекста с установлением новых иерархических отношений между ними.

1.5. Методика определения эпистемической оценки

Описываемая методика предполагает наличие текста с заранее размеченными целевыми пропозициями. Для определения ЭО целевой пропозиции следует перейти от полной иерархической синтаксической модели (описывающей поверхностно-синтаксический уровень) к модели, которая отражала бы иерархию предикативных вершин в полипредикативном предложении на уровне глубинного синтаксиса. Для этого в изначальной иерархической синтаксической модели выбираются узлы, соответствующие предикативным вершинам, и создается новая древовидная структура, в которой эти узлы связаны напрямую (включающий предикат соответствует узлу на более высоком уровне, включенный предикат — узлу на более низком уровне).

Для работы в рамках этой модели требуются следующие понятия:

Исходное дерево — синтаксическая модель предложения, построенная анализатором MaltParser.

Целевой узел соответствует предикатной вершине целевой пропозиции.

Опорные узлы — узлы исходного дерева, соответствующие предикатным вершинам в предложении.

Сокращенное дерево опорных узлов — древовидная структура, образованная из опорных узлов путем сокращения исходного дерева.

Подчиняющий узел — узел сокращенного дерева, имеющий потомков.

Проецируемое значение ЭО — значение ЭО, которое подчиняющий узел проецирует на своего непосредственного потомка, устанавливается путем обращения к словарю.

²Перечень этих синтаксических отношений был выработан в Лаборатории компьютерной лингвистики Института проблем передачи информации РАН. См. www.ruscorpora.ru

Эпистемическая оценка узла — сокращенная запись, которая подразумевает «ЭО предиката, соответствующего узлу».

Собственная эпистемическая оценка узла — атрибут, в который записывается предварительная ЭО узла. Значение собственной ЭО устанавливается на основании морфологических характеристик фрагмента, на который ссылается узел, и определенных зависимых от него узлов, в числе которых могут быть, например, вводные слова, маркирующие сниженную ЭО, условные конструкции («при условии...», «если...») и т.п. Собственная ЭО принимает значение «фактуально», если узел не имеет зависимых узлов, влияющих на значение ЭО (как перечислены выше) и если фрагмент, на который ссылается узел, содержит личную форму глагола в индикативе.

Окончательная эпистемическая оценка узла — атрибут, значение которого устанавливается с учетом собственной ЭО и проецируемой ЭО узлом более высокого уровня.

2. Экспериментальные результаты

2.1. Построение сокращенного дерева

Прежде чем непосредственно перейти к установлению значения ЭО для целевой пропозиции, следует провести подготовительный этап: анализ поверхностно-синтаксических отношений. Целью этого анализа является обнаружение в исходном дереве релевантных предикативных и полупредикативных конструкций (точнее, их предикативную вершину) и установить, в каких иерархических связях между собой они находятся.

В терминах описываемой методики это означает, что требуется проанализировать иерархическую синтаксическую модель, полученную в результате работы MaltParser, выделить в предложении опорные узлы и установить связи между ними. На рисунке 2 приведен фрагмент исходной синтаксической модели в системе анализа текстов ИСИДА-Т [14] для предложения: «*Сообщалось, что в новогоднюю ночь полиция задержала 150 человек.*».

Сообщалось	что	в	новогоднюю	ночь	полиция	задержала	
SyntaxNode [35] Generator leafType Morpho Morpho Сообщалось [25]	SyntaxNode [36] Generator leafType Morpho Morpho [24] Parent PhraseNode Сообщалось [25] leafType Phrase	SyntaxNode [37] Generator leafType Morpho Morpho 'что' [23] Parent PhraseNode Сообщалось [25] leafType Phrase	SyntaxNode [38] Generator leafType Morpho Morpho 'в' [22] Parent PhraseNode 'что' [23] leafType Phrase	SyntaxNode [39] Generator leafType Morpho Morpho новогодней [27] Parent PhraseNode 'что' [23] leafType Phrase	SyntaxNode [40] Generator leafType Morpho Morpho 'ночь' [26] Parent PhraseNode 'в' [22] leafType Phrase	SyntaxNode [41] Generator leafType Morpho Morpho полиция [28] Parent PhraseNode 'задержала' [42] leafType Phrase	SyntaxNode [42] Generator leafType Morpho Morpho задержала [30] Morpho Parent PhraseNode 'что' [23] leafType Phrase

Рисунок 2. Фрагмент исходной иерархической синтаксической модели

Опорные узлы выделяются попарно, для каждой пары должны соблюдаться следующие условия:

- каждый из двух узлов имеет заданное значение атрибутов («часть речи», «полярность» и др);
- один из узлов входит в число потомков другого узла (непосредственно или опосредованно), атрибут «синтаксическая роль» имеет заранее заданное значение (в том случае, если узлы связаны опосредованно, через другие узлы, эти узлы-посредники также имеют заданное значение атрибута «синтаксическая роль»);
- узел, находящийся на более высоком уровне в иерархической синтаксической модели, ссылается на фрагмент текста, каноническая (словарная) форма которого входит в «Словарь подчиняющих предикатов»³.

В общей сложности описано 22 таких условия.

На рисунке 3 можно увидеть пример условной части правила, которое находит два таких узла. В рассмотренном выше примере это

```
{И (
{ЗВЕНО_СВЯЗИ (Node)}
{ЗВЕНО_СВЯЗИ ( {РОДИТЕЛЬ({РОДИТЕЛЬ (Node)}) } ) }
{АТРИБУТ_УЗЛА_РАВЕН (Node "parser_roleподч-союз")}
{АТРИБУТ_УЗЛА_РАВЕН ({РОДИТЕЛЬ (Node)} "parser_canonicalчто")}
{АТРИБУТ_УЗЛА_РАВЕН ({РОДИТЕЛЬ (Node)} "parser_role1-компл")}
)}
```

Рисунок 3. Пример условной части правила

правило применимо для узла «задержала». Для него выполняются все условия: сам узел удовлетворяет условиям функции (эта функция проверяет, к какой части речи относится слово), атрибут узла «parser_role» имеет значение «подч-союз», для его родителя атрибут «parser_canonical» (в этом атрибуте хранится каноническая форма слова) равен «что», а атрибут «parser_role» равен «1-компл», родитель его родителя («сообщалось») также удовлетворяет условиям функции.

В командной части (не приведена на рисунке) сам узел и родитель его родителя («задержала» и «сообщалось») сохраняются в пространство контекста, при этом между ними устанавливаются новые иерархические отношения. В итоге после обработки всего предложения в пространстве контекста сохранено сокращенное дерево опорных

³со "Словарем подчиняющих предикатов" и другими дополнительными материалами к статье можно ознакомиться по адресу <https://github.com/Maienblume/Factuality-status>

узлов, вершины которого соответствуют предикативным вершинам. На рисунке 4 приведен пример визуализации такого дерева.

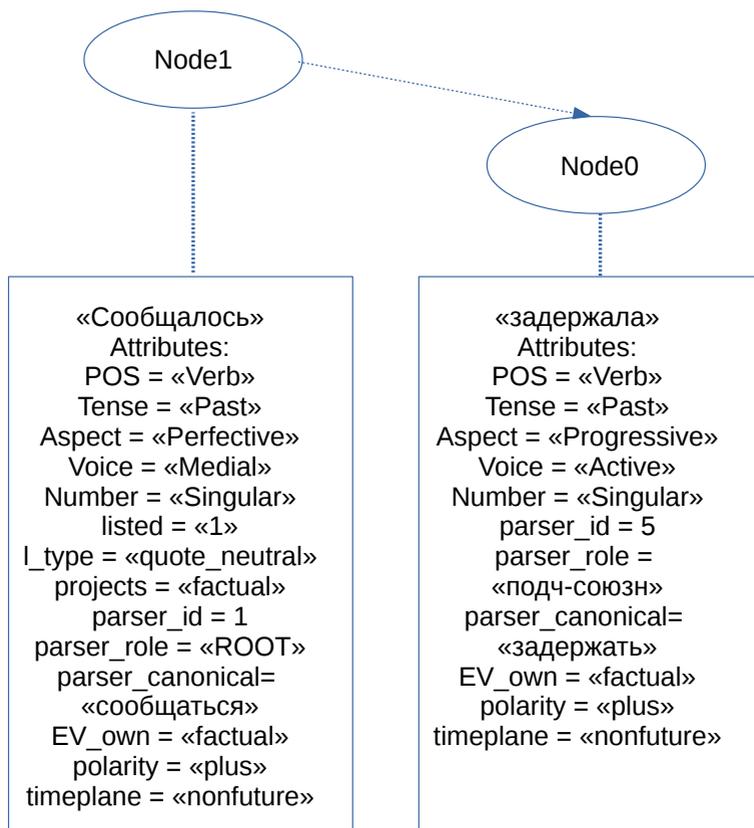


Рисунок 4. Сокращенное дерево опорных узлов

Одновременно с этим узлам сокращенного дерева присваиваются атрибуты, которые будут использоваться впоследствии. Самым важным из них является «собственная эпистемическая оценка узла».

2.2. Присвоение значений эпистемической оценки

После создания дерева можно переходить к основному этапу: присвоению значений ЭО опорным узлам. На этом этапе правила обрабатывают только сокращенное дерево опорных узлов. Для каждого узла требуется установить окончательную ЭО. В условной части

правил этого этапа проверяется:

- собственная ЭО узла (определена ранее);
- собственная ЭО и временной план родителей узла — способность подчиняющих узлов проецировать значение ЭО на своего потомка зависит от значений обоих этих атрибутов или одного из них;
- проецируемое значение непосредственного родителя узла;
- проецируемое значение всех родителей узла, кроме непосредственного. Если все вышестоящие родители являются нейтральными, то есть не оказывают никакого влияния на значения ЭО подчиненного узла («дерево нейтрально»), то в качестве окончательного ЭО узла устанавливается то значение, которое проецирует его непосредственный родитель. Если хотя бы один родитель (кроме непосредственного родителя узла) проецирует значения «нефактуальная ЭО» или «контрфактуальная ЭО» («дерево не нейтрально»), то окончательная ЭО рассматриваемого узла принимает значение «не определено».

В итоге каждый узел может получить одно из следующих значений окончательной ЭО (для каждого значения приведен набор условий, которые должны соблюдаться для присвоения этого значения узлу):

Фактуально:

Вариант 1:

- предварительная ЭО узла была определена как «фактуальная» и
- непосредственный родитель проецирует значение «абсолютно-фактуально».

Вариант 2:

- предварительная ЭО узла была определена как «фактуальная» и
- «дерево нейтрально»

Контрфактуально:

- предварительная ЭО узла была определена как «фактуальная»,
- непосредственный родитель проецирует значение «контрфактуально»,
- «дерево нейтрально»

Нефактуально:

- предварительная ЭО определена как «нефактуально» (например, при наличии вводных слов, снижающих ЭО, таких как «возможно») или непосредственный родитель проецирует значение «контрфактуально»

- «дерево нейтрально».

Не определено:

- непосредственный родитель проецирует значение «неопределено» (это предикаты, пропозиционный актанта которых имеет ирреальную модальность — например, «приказать», «требовать», «хотеть»...) или
- «дерево не нейтрально».

Присвоенное значение (наряду с временным планом) фиксируются в аннотации в системе ИСИДА-Т. В результате работы всех правил для каждого опорного узла создается такая аннотация (рисунок 5).

Сообщалось * , * что * в * новогоднюю * ночь * полиция * задержала * 150 * человек * .	
Factuality_profile [Z1] Epistemic_value factuel Generator SPSL Polarity plus Timeplane nonfuture	Factuality_profile [Z2] Epistemic_value factuel Generator SPSL Polarity plus Timeplane nonfuture

Рисунок 5. Аннотации с фактуальным статусом для опорных узлов

2.3. Примеры разметки

В данном разделе приведены примеры разметки, произведенной описанными правилами. Каждый опорный узел получает собственную аннотацию (разный цвет аннотаций объясняется особенностями работы системы ИСИДА-Т и не несет никакой смысловой нагрузки).

Как видно из иллюстраций, опорные узлы могут находиться как в пределах одного простого предложения (рисунки 6, 7), так и в разных частях сложного предложения с разным типом придаточных (рисунки 8, 9, 10). В примерах можно видеть разные значения атрибута «Epistemic_value» (окончательная эпистемическая оценка узла).

О * возможной * поездке * российского * лидера * проинформировал * заместитель * губернатора * Ростовской * области * Вадим * Артемов * .	
Factuality_profile [Z0] Epistemic_value nonfactual Generator SPSL	Factuality_profile [Z1] Epistemic_value factuel Generator SPSL Polarity plus Timeplane nonfuture

Рисунок 6. «О возможной поездке российского лидера проинформировал заместитель губернатора Ростовской области Вадим Артемов»

Большинство опорных узлов являются глагольными формами, однако в определенных случаях в качестве опорного может быть выбрано и существительное: «поездка», «отставка» (рисунки 6, 7).

Я	был	просто	поражен	и	шокирован	отставкой	Поковича
	Factuality_profile [67] Epistemic_value factual Generator SPSSL Polarity plus Timeplane nonfuture		Factuality_profile [68] Epistemic_value factual Generator SPSSL Polarity plus Timeplane nonfuture		Factuality_profile [69] Epistemic_value factual Generator SPSSL Polarity plus Timeplane nonfuture		Factuality_profile [70] Epistemic_value factual Generator SPSSL

Рисунок 7. «Я был просто поражен и шокирован отставкой Поковича»

Эксперты	сомневаются	что	Папа Римский	приедет	в	Беларусь
	Factuality_profile [18] Epistemic_value factual Generator SPSSL Polarity plus Timeplane nonfuture			Factuality_profile [19] Epistemic_value nonfactual Generator SPSSL Polarity plus Timeplane future		

Рисунок 8. «Эксперты сомневаются, что Папа Римский приедет в Беларусь»

Брэдли	абсолютно	отрицает	что	совершил	преступление
		Factuality_profile [18] Epistemic_value factual Generator SPSSL Polarity plus Timeplane nonfuture		Factuality_profile [19] Epistemic_value counterfactual Generator SPSSL Polarity plus Timeplane nonfuture	

Рисунок 9. «Брэдли абсолютно отрицает, что совершил преступление»

Шнуров	рассказал	почему	не	везет	группу	на	Украину
	Factuality_profile [20] Epistemic_value factual Generator SPSSL Polarity plus Timeplane nonfuture			Factuality_profile [21] Epistemic_value factual Generator SPSSL Polarity minus Timeplane nonfuture			

Рисунок 10. «Шнуров рассказал, почему не везет группу на Украину»

Опорные узлы-существительные не имеют атрибутов «полярность» и «временной план». Теоретически «преступление» на рисунке 9 также может быть размечено как опорный узел, однако на данный момент описанные правила в первую очередь ориентируются на глагольные формы.

2.4. Оценка эффективности правил

Эффективность работы правил проверялась на материале тренировочного и тестового множества. Для подсчета статистики использовались предложения, в которых за «элемент разметки» принимались предикаты событийных пропозиций, выраженные преимущественно

глагольными формами. Событийные существительные принимались во внимание лишь в тех случаях, где они подчинены заранее описанным предикатам из «Списка подчиняющих предикатов». Таким образом, если в предложении присутствовало событийное существительное, не связанное с таким предикатом, отсутствие аннотации для него не считалось за ошибку.

В тренировочном множестве правилами были обработаны 100 предложений, содержащих 291 элемент разметки. Предложения для этого множества подобраны так, чтобы охватить максимально возможное количество разнообразных факторов, влияющих на эпистемическую оценку предиката (наличие вводных слов и условных конструкций, разные типы подчиняющих предикатов, подчинительные и сочинительные связи между предикатами и т. п.), т. е. множество репрезентативно. Тестовое множество состоит из случайно выбранных новостных текстов СМИ, содержащих 104 элемента разметки, из которых 64 «целевых». Результаты приведены в таблице 1.

Таблица 1. Оценка эффективности работы правил

	Тренировочное множество	Тестовое множество
точность	0,94	0,84
полнота	0,86	0,89
F-мера	0,898	0,87

Заключение

В статье предложена методика, которая позволяет установить значение всех компонентов фактуального статуса для каждой событийной пропозиции в предложении: эпистемическую оценку (основной компонент), временной план и отношение автора к цитируемому источнику. Компонент «эпистемическая оценка» (ЭО) может принимать значения: «фактуально», «контрфактуально», «нефактуально» и «не определено».

Согласно данной методике, для установления значения «эпистемическая оценка» пропозиции требуется:

- (1) представить полипредикативное предложение в виде «дерева опорных узлов», отражающего иерархию пропозиционных предикатов в предложении;
- (2) для каждого опорного узла определить «предварительную ЭО» на основании собственных морфологических характеристик предиката и его зависимых узлов;
- (3) для каждого подчиняющего опорного узла определить тип его влияния на непосредственного потомка (проецируемое значение ЭО) с помощью «Словаря подчиняющих узлов»;
- (4) определить окончательную ЭО каждого узла на основании его предварительной ЭО и значения ЭО, проецируемого вышестоящим опорным узлом.

Дальнейшие исследования должны вестись в следующих направлениях:

1. Расширение количества контекстов, которые опознаются правилами как «опорные узлы» и используются для построения сокращенного дерева опорных узлов. В настоящий момент как опорные узлы опознаются в первую очередь глагольные формы, следует уделить более пристальное внимание именным способам референции к пропозиции (таким как отглагольное существительное).
2. Для расширенного количества контекстов потребуется и расширенный список условий, при выполнении которых эти контексты оказывают влияние на фактуальный статус подчиненной пропозиции (например, для пропозиции, выраженной именем способом, таким условием является участие в определенных предложных группах).
3. Уточнение возможных значений «эпистемическая оценка», особенно значения «нефактуальная ЭО». Для более адекватного представления фактуального статуса пропозиции следует ввести градацию нефактуальной ЭО (от низкой степени эпистемической ответственности до более высокой).

Благодарности. Авторы выражают признательность Е. А. Сулеймановой и И. В. Трофимову за внесенные замечания.

Список литературы

- [1] R. Saurí, J. Pustejovsky. “Factbank: a corpus annotated with event factuality”, *Language resources and evaluation*, **43**:3 (2009), pp. 227–268. 
  ^{24, 27}
- [2] A.-L. Minard, M. Speranza, R. Urizar, B. Altuna, M. van Erp, A. Schoen, C. van-Son. “MEANTIME, the NewsReader multilingual event and time corpus”, *Proceedings of LREC2016* (May 2016, Portorož, Slovenia), ELRA, 2016, pp. 4417–4422.  ²⁴
- [3] K. Lee, Y. Artzi, Y. Choi, L. Zettlemoyer. “Event detection and factuality assessment with non-expert supervision”, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP’2015* (September 17–21, 2015, Lisbon, Portugal), ACL, 2015, ISBN 978-1-941643-32-7, pp. 1643–1648.  ²⁴
- [4] G. Stanovsky, J. ECKLE-KOHLER, E. Puzikov, I. Dagan, I. Gurevych. “Integrating deep linguistic features in factuality prediction over unified datasets”, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Short Papers. V. 2* (July 30–August 4, 2017, Vancouver, Canada), ACL, pp. 352–357.   ²⁴
- [5] R. Saurí, J. Pustejovsky. “Are you sure that this happened? Assessing the factuality degree of events in text”, *Computational Linguistics*, **38**:2 (2012), pp. 261–299.   ²⁴
- [6] A. Lotan, A. Stern, I. Dagan. “Truth-teller: Annotating predicate truth”, *Proceedings of NAACL-HLT 2013* (9–14 June 2013, Atlanta, Georgia), ACL, pp. 752–757.  ²⁴
- [7] Е. В. Падучева. «Снятая утвердительность и неверидикативность» (Бекасово, 4–8 июня 2014 г.), Компьютерная лингвистика и интеллектуальные технологии, т. **13(20)**, По материалам ежегодной Международной конференции «Диалог», РГГУ, М., 2014, с. 489–505.  ²⁶
- [8] Е. В. Падучева. *О семантике синтаксиса. Материалы к трансформационной грамматике русского языка*, Либроком, М., 2009, ISBN 978-5-397-06844-4. ²⁶
- [9] Русграм, <http://rusgram.ru/>. ²⁶
- [10] Е. В. Падучева. *Эгоцентрические единицы языка*, 2-е изд., Издательский дом ЯСК, М., 2019, ISBN 978-5-907117-23-5, 440 с. ²⁷
- [11] F. R. Palmer, *Mood and modality*, Cambridge Textbooks in Linguistics, 2nd ed., Cambridge University Press, Cambridge, 2001, ISBN 9781139167178.  ²⁷
- [12] А. А. Зализняк. ««Знание» и «мнение» в семантике предикатов внутреннего состояния», *Коммуникативные аспекты исследования языка*, Изд-во ин-та языкознания АН СССР, М., 1986, с. 4–15. ^{27, 32}
- [13] Н. Д. Арутюнова. «Будущее в языке», *Логический анализ языка. Лингвофутуризм. Взгляд языка в будущее*, Индрик, М., 2011, ISBN 978-5-91674-130-8, с. 65–80. ²⁸

- [14] Д. А. Кормалев, Е. П. Куршев, Е. А. Сулейманова, И. В. Трофимов. «Извлечение информации из текста в системе ИСИДА-Т», *Труды 11й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии электронные коллекции»*, RCDL'2009 (17–21 сентября 2009 г., Петрозаводск, Россия), 2009, с. 247–253. ↑₃₅

Поступила в редакцию 29.11.2019

Переработана 22.03.2020

Опубликована 05.05.2020

Рекомендовал к публикации

к.т.н. Е. П. Куршев

Пример ссылки на эту публикацию:

С. Р. Момот, И. Н. Воздвиженский. «Оценка фактуальности для пропозиции в синтаксически подчиненном положении». *Программные системы: теория и приложения*, 2020, **11**:2(45), с. 23–46.

 10.25209/2079-3316-2020-11-2-23-46

 http://psta.psiras.ru/read/psta2020_2_23-46.pdf

Об авторах:



Седа Рубеновна Момот

Младший научный сотрудник Исследовательского центра искусственного интеллекта ИПС им. А. К. Айламазяна РАН. Область научных интересов: компьютерная лингвистика, автоматическая обработка естественного языка, извлечение событийной информации.

 0000-0002-6097-6545

e-mail: seda.egikian@gmail.com



Илья Николаевич Воздвиженский

младший научный сотрудник Исследовательского центра искусственного интеллекта ИПС им. А. К. Айламазяна РАН, кандидат технических наук, один из разработчиков систем автоматической обработки естественного языка, в том числе разработчик языка правил для анализа синтаксических отношений.

 0000-0001-8959-3460

e-mail: vozdvin@mail.ru

CSCSTI 28.23.13,16.31.21
UDC 004.89:004.912

Seda R. Momot, Il'ya N. Vozdvizhenskiy. *The assessment of the factuality of the syntactically subordinated proposition.*

ABSTRACT. The article presents a method to define the factuality status of a proposition in the sphere of the natural language processing. A particular emphasis is placed upon syntactically subordinated propositions. The term "factuality status" is defined by introduction of its three components, the core component is the so-called "epistemic value". The essence of the method is the hierarchical alignment of the predicates of the constructions in a polipredicative sentence. An epistemic value is assigned to every predicate by a set of rules and a dictionary.

Key words and phrases: factuality, factuality status, epistemic value, time plane, hypotaxical constructions, predicate, proposition.

2010 *Mathematics Subject Classification:* 68T50; 91F20

References

- [1] R. Saurí, J. Pustejovsky. "Factbank: a corpus annotated with event factuality", *Language resources and evaluation*, **43:3** (2009), pp. 227–268.  [↑_{24,27}](#)
- [2] A.-L. Minard, M. Speranza, R. Urizar, B. Altuna, M. van Erp, A. Schoen, C. van-Son. "MEANTIME, the NewsReader multilingual event and time corpus", *Proceedings of LREC2016* (May 2016, Portorož, Slovenia), ELRA, 2016, pp. 4417–4422.  [↑₂₄](#)
- [3] K. Lee, Y. Artzi, Y. Choi, L. Zettlemoyer. "Event detection and factuality assessment with non-expert supervision", *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP'2015* (September 17–21, 2015, Lisbon, Portugal), ACL, 2015, ISBN 978-1-941643-32-7, pp. 1643–1648.  [↑₂₄](#)
- [4] G. Stanovsky, J. Eckle-Kohler, E. Puzikov, I. Dagan, I. Gurevych. "Integrating deep linguistic features in factuality prediction over unified datasets", *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Short Papers. V. 2* (July 30–August 4, 2017, Vancouver, Canada), ACL, pp. 352–357.   [↑₂₄](#)
- [5] R. Saurí, J. Pustejovsky. "Are you sure that this happened? Assessing the factuality degree of events in text", *Computational Linguistics*, **38:2** (2012), pp. 261–299.   [↑₂₄](#)
- [6] A. Lotan, A. Stern, I. Dagan. "Truth-teller: Annotating predicate truth", *Proceedings of NAACL-HLT 2013* (9–14 June 2013, Atlanta, Georgia), ACL, pp. 752–757.  [↑₂₄](#)
- [7] Ye. V. Paducheva. "Suspended assertion and nonveridicality" (Bekasovo, June 4–8, 2014), *Computational Linguistics and Intellectual Technologies*, vol. **13(20)**, papers from the Annual conference "Dialogue", RGGU, M., 2014, pp. 489–505 (in Russian).  [↑₂₆](#)

- [8] Ye. V. Paducheva. *On Semantics of syntax. Materials to the transormational grammar of the Russian language*, Librokom, M., 2009, ISBN 978-5-397-06844-4 (in Russian).[↑]₂₆
- [9] Rusgram, <http://rusgram.ru/>.[↑]₂₆
- [10] Ye. V. Paducheva. *Egocentric language units*, 2-ye izd., Izdatel'skiy dom YaSK, M., 2019, ISBN 978-5-907117-23-5 (in Russian), 440 pp.[↑]₂₇
- [11] F. R. Palmer, *Mood and modality*, Cambridge Textbooks in Linguistics, 2nd ed., Cambridge University Press, Cambridge, 2001, ISBN 9781139167178. [↑]₂₇
- [12] A. A. Zaliznyak. ““Knowledge” and “Opinion” in the semantics of predicates of the internal state”, *Kommunikativnyye aspekty issledovaniya yazyka*, Izd-vo in-ta yazykoznaniya AN SSSR, M., 1986, pp. 4–15 (in Russian).[↑]_{27, 32}
- [13] N. D. Arutyunova. “Future in the Language”, *Logicheskii analiz yazyka. Lingvofuturizm. Vzgl'yad yazyka v budushcheye*, Indrik, M., 2011, ISBN 978-5-91674-130-8, pp. 65–80 (in Russian).[↑]₂₈
- [14] D. A. Kormalev, Ye. P. Kurshev, Ye. A. Suleymanova, I. V. Trofimov. “Information extraction in ISIDA-T system”, *Proceedings of the 11th conference “Digital Libraries: Advanced Methods and Technologies”*, RCDL'2009 (17–21 sentyabrya 2009 g., Petrozavodsk, Rossiya), 2009, pp. 247–253 (in Russian).[↑]₃₅

Sample citation of this publication:

Seda R. Momot, Il'ya N. Vozdvizhenskiy. “The assessment of the factuality of the syntactically subordinated proposition”. *Program Systems: Theory and Applications*, 2020, **11**:2(45), pp. 23–46. (In Russian).  10.25209/2079-3316-2020-11-2-23-46

 http://psta.pspiras.ru/read/psta2020_2_23-46.pdf