

23–25 апреля 2025 г.

Towards the task of factuality assessment

Elena A. Suleymanova

A. K. Aylamazyan Program Systems
Institute of RAS,
Pereslavl-Zalessky, Russia
yes2helen@gmail.com

Seda R. Momot

A. K. Aylamazyan Program Systems
Institute of RAS,
Pereslavl-Zalessky, Russia
seda.egikian@gmail.com

Natalia A. Vlasova

A. K. Aylamazyan Program Systems
Institute of RAS,
Pereslavl-Zalessky, Russia
nathalie.vlassova@gmail.com

Ilya N. Vozdvizhensky

A. K. Aylamazyan Program Systems
Institute of RAS,
Pereslavl-Zalessky, Russia
vozdvin@yandex.ru

Abstract

Factuality concerns the extent to which the propositional content of a sentence conforms with the real world, according to the speaker. Factuality assessment is of practical interest for those NLP applications that have to do with textual information analysis: information extraction, information retrieval, text summarization, question-answering systems.

Existing practical approaches address the problem of factuality assessment using the “quantitative” measures of certainty and probability and thus have limited possibilities for annotating real texts. There is a need for a more appropriate model of factuality that could serve as the basis of the annotation scheme.

We suggest a model of factuality that makes use of speaker's cognitive attitudes as one of the parameters for discriminating between degrees of factuality. We also present a dataset that was manually annotated with factuality values in terms of the model.

Keywords: Keywords: factuality assessment; factuality model; annotating factuality in text; dataset annotated with factuality

DOI: 10.28995/2075-7182-2025-23-XX-XX

К задаче оценки субъективной достоверности

Сулейманова Е. А.

Институт программных систем
им. А. К. Айламазяна РАН
Переславль-Залесский, Россия
yes2helen@gmail.com

Момот С. Р.

Институт программных систем
им. А. К. Айламазяна РАН
Переславль-Залесский, Россия
seda.egikian@gmail.com

Власова Н. А.

Институт программных систем
им. А. К. Айламазяна РАН
Переславль-Залесский, Россия
nathalie.vlassova@gmail.com

Воздвиженский И. Н.

Институт программных систем
им. А. К. Айламазяна РАН
Переславль-Залесский, Россия
vozdvin@yandex.ru

Аннотация

Субъективная достоверность, или авторская квалификация сообщаемого по линии соответствия действительности, представляет практический интерес для тех приложений автоматической обработки текста, которые связаны с содержательным анализом информации: извлечение информации из неструктурированных источников, информационный поиск, аннотирование, реферирование, вопросно-ответные системы.

Существующие практические подходы к оценке субъективной достоверности оперируют исключительно «количественными» признаками уверенности и вероятности, что ограничивает возможности их применения для аннотирования реальных текстов. Актуальной остается задача построения более адекватной модели субъективной достоверности, которая могла бы лечь в основу аннотационной схемы.

В статье представлена модель субъективной достоверности, использующая в качестве одного из оснований для дифференциации значений противопоставление когнитивных установок говорящего. В терминах модели размечен датасет.

Ключевые слова: оценка субъективной достоверности; модель субъективной достоверности; аннотирование субъективной достоверности в тексте; датасет с разметкой субъективной достоверности

1 Введение

Субъективная достоверность (далее «СД») — аспект текста, отвечающий за квалификацию говорящим сообщаемого с точки зрения соответствия действительности.

Возможности современных нейросетевых технологий позволяют предположить, что решение задачи компьютерного анализа текста с т. зр. СД может быть в значительной степени сведено к подготовке релевантных задаче обучающих данных. Для аннотирования таких данных необходима теоретическая модель, которая, на наш взгляд, должна (1) охватывать всё разнообразие языковых средств, образующих функционально-семантическое поле СД, и (2) отвечать языковой интуиции.

Задача, подобная аннотированию СД, возникла за рубежом в русле работ по извлечению событий и известна под названием «оценка фактуальности»¹ (подробнее см. в разделе 2). В большинстве исследований используется шкала «certain-probable-possible», предложенная R. Sauri и J. Pustejovsky в 2008-2009 годах для корпуса FactBank [Sauri 2009]. Пытаясь применить эту шкалу на практике, мы регулярно сталкивались с одной и той же проблемой — отсутствием подходящего значения². Идеи отечественных лингвистов (раздел 3) укрепили нас в мнении, что для описания значений СД «количественных» признаков уверенности и вероятности недостаточно. Рассмотрим пары примеров:

(1а) *Я считаю, что это сделал он.*

(1б) *Я считаю, что это следовало сделать.*

(2а) *Мне кажется, он пришел за объяснением.*

(2б) *Мне кажется, пришло время об этом поговорить.*

(3а) *Едва ли его за это похвалили.*

(3б) *Едва ли это заслуживает похвалы.*

(4а) *Я не думаю, что руководству известны причины произошедшего.*

(4б) *Я не думаю, что об этом лучше молчать.*

(5а) *Он точно дома. [Я видел свет в окне] Ср. также *Он явно/определенно дома*.*

(5б) *Он точно дома! [Не сомневайся, я только что от него] Ср. **Он явно/определенно дома*.*

(6а) [Я не был на собрании] ***Наверное/вероятно, выбрали Петрова***. [Он самая подходящая кандидатура] Ср. **Вроде/кажется, выбрали Петрова*.

(6б) [Да, я был на собрании.] ***Вроде/кажется, выбрали Петрова***. Ср. **Наверное/вероятно, выбрали Петрова*).

Сразу заметим, что степень «уверенности», выраженная показателями СД (они выделены жирным шрифтом), в каждой паре «а» и «б», очевидно, одинакова.

¹ Английский термин factuality следует, по-видимому, признать самым близким аналогом термина «субъективная достоверность».

² Из-за этого обстоятельства мы вынуждены были отказаться от сравнения нашего подхода с «мейнстримом» на датасете: непонятно, каким образом при подсчете коэффициента согласия учесть единодушное «затрудняюсь».

Во все парах пример «а» описывает случай, когда действительное положение дел (имеет ли место ситуация, выраженная подчеркнутым фрагментом) говорящему не известно и он высказывает предположение об этом, основываясь на косвенных данных. Во всех парах пример «а» допускает естественную интерпретацию в терминах вероятности, чего нельзя сказать о примерах «б».

В примерах (1б)-(4б) подчеркнутый фрагмент обозначает не ситуацию реального мира, а собственный взгляд говорящего, его субъективную точку зрения. В этих случаях показатель СД выражает степень уверенности, категоричности мнения; усмотреть в них вероятностную оценку либо нельзя, либо можно с очень большой натяжкой.

В примере (5б) говорящий доподлинно знает то, о чем сообщает. Наречие «точно» в данном случае не является маркером СД, это дискурсивный маркер настойчивой утвердительности (в ответ на сомнения собеседника; заметим, что примеры (5а) и (5б) различаются и интонационно).

Разница между (6а) и (6б) не так очевидна. Она в том, что в (6а) говорящий не знает, а только строит догадки. А в (6б) нет ничего похожего на предположение — это знание, которое говорящий осторожно квалифицирует как не совсем надежное. Различие в значениях СД между (6а) и (6б) подтверждается невозможностью взаимной замены в них показателей СД.

В следующем разделе мы попытаемся описать приведенные примеры в терминах подходов state-of-the-art.

Подход, который предлагается в настоящей статье, основывается на следующих положениях.

1. Значения СД могут иметь различную природу. Это относится как к случаям подчеркнутой достоверности (присутствие уверенности), так и к случаям проблематичной достоверности (отсутствие уверенности). И в тех, и в других случаях значение может быть вероятностным, а может и не быть таковым.
2. Природа (качественный тип) значения СД определяется взаимодействием трех факторов:
 - выбор говорящим маркера СД для оформления пропозиционального содержания,
 - характер самой пропозиции,
 - контекст (в т. ч. прагматический — условия речевого акта, в широком смысле).
3. Маркеры СД часто неоднозначны: один и тот же маркер способен обслуживать разные качественные типы значений СД. Для разрешения неоднозначности маркера важно принимать во внимание характер пропозиции и контекст.

Целью настоящей работы было проверить работоспособность подхода в целом. Построенная для этого модель в ее текущей версии имеет ряд ограничений:

- а) рассматривается квалификация пропозиции только одним субъектом — говорящим³;
- б) рассматриваются только случаи, в которых оценка СД выражена лексическими средствами;
- в) принятые по некоторым сложным и спорным случаям решения не могут считаться окончательными.

Модель была испытана на 510 реальных текстовых примерах с последующим подсчетом коэффициента согласия аннотаторов (каппа Флейса). Результаты разметки примеров в соответствии с предложенной моделью оформлены как общедоступный датасет.

2 Субъективная достоверность в компьютерной лингвистике

В компьютерной лингвистике анализом явлений, в той или иной степени имеющих отношение к СД, занимаются несколько направлений: оценка фактуальности (factuality assessment), аннотирование мнений (belief annotation and tagging), оценка достоверности/уверенности (certainty evaluation), анализ субъективности (subjectivity analysis) и установки говорящего (speaker's attitude), выявление субъективности (subjectivity detection). Эти исследования можно условно разделить на две группы.

Первая группа представлена работами по оценке фактуальности [Saurí, Pustejovsky 2012], [Lee 2015], [Lima 2020], [Gupta 2022], [Li 2024], [Rovera 2025]. Для этого направления интерес представляет то, как «событие» (event) соотносится с действительностью, с точки зрения субъекта. Оценка (значение фактуальности) представляет собой комбинацию признака «полярность» («+»),

³ Случаям передачи чужой речи (чужой пропозиции) посвящено отдельное исследование.

«→») с некоторым количественным признаком *certainly* (достоверность, она же уверенность). Достоверные/уверенные случаи (*certain, CT*) противопоставляются недостоверным (*uncertain*), для оценки которых используются разные степени вероятности. В классической версии их две — *probable (PR)*, и *possible (PS)*. Иногда вместо этой шкалы используется непрерывная числовая шкала $[-3;+3]$ [Lee 2015]. Примечательно, что к *certain* авторы концепции относят как имплицитно достоверные случаи (простые утверждения, презумпции, импликации), так и случаи с лексическими маркерами уверенности (*certain, sure*) и модальными показателями эпистемической необходимости (*must, have to*). Иными словами, «событие» ‘он дома’ в примерах *Он дома; Они (не) знают, что он дома; Я уверен, что он дома* и *Должно быть, он дома* получает одинаковое значение фактуальности (*CT+*)⁴.

Что касается второй группы исследований, то в центре их внимания — субъективность как таковая. Это либо разделение всех пропозиций (предикатов, утверждений) на уверенное и неуверенное мнение (убеждение) — *committed belief* и *non-committed belief* [Diab 2009], [Prabkharan 2015], [García 2020], либо разделение высказываний на «субъективные» и «объективные» [Antici 2024]. Вопросы соотносительности самой пропозиции с действительностью в подобных работах не затрагиваются, что не позволяет отнести используемые ими модели к полноценным моделям СД.

Оценка примеров из раздела «Введение» на шкале фактуальности и шкале оценки мнений выглядит следующим образом (вопросом помечены значения, которые приписаны за неимением более подходящих):

Пример	Значение по шкале [Saurí 2009]	Значение по шкале [Diab 2009]
1a	CT+?	committed belief
1б	CT+	committed belief
2a	PR+	non-committed belief
2б	PR+?	non-committed belief
3a	PR-	non-committed belief
3б	PR-?	non-committed belief
4a	PR-	non-committed belief
4б	PR-?	non-committed belief
5a	CT+	committed belief
5б	CT+?	committed belief
6a	PR+	non-committed belief
6б	PR+?	non-committed belief

Таблица 1: Разметка примеров по шкале фактуальности и шкале мнений (belief)

Легко видеть, что примеры «а» и «б» во всех парах получают одинаковые значения.

3 Теоретические предпосылки для альтернативного подхода к оценке субъективной достоверности

В отечественных теоретических работах по языкознанию можно найти много ценных наблюдений на интересующую нас тему.

⁴ Значение фактуальности всегда привязано к «источнику» (source) — в данном случае это author.

[Яковлева 1994] считает, что при классификации показателей достоверности, кроме количественных признаков, нужно учитывать качество информации: характерная, полученная субъектом непосредственно (*кажется, как будто, вроде*) и нехарактерная, полученная на основе логического вывода (*наверно, видимо*).

[Булыгина, Шмелёв 1997] выделяют на этом основании два типа показателей недостоверности: показатели гипотетичности (у говорящего нет информации об истинности, он высказывает предположение, основываясь на логическом выводе или интуитивной догадке) и квазиассертивы (говорящий имеет непосредственный доступ к информации, но по каким-то причинам не совсем уверен в ее достоверности).

[Шмелёва 1984] усматривает в выражении семантики СД взаимодействие модальных категорий персуазивности (квалификация сообщаемого с т. зр. соответствия действительности) и авторизации (квалификация с т. зр. источников или способов получения информации).

[Дмитровская 1988], [Зализняк 1992] в сфере «знание-мнение» различают три типа установок: знание о верифицируемой пропозиции (знание), мнение о верифицируемой пропозиции (мнение-предположение) и мнение о неверифицируемой пропозиции (мнение-оценка). Мнение-оценка, как и знание, включает истинностную оценку подчиненной пропозиции, тогда как мнение-предположение — вероятностную. Мнение-оценку можно назвать «субъективным знанием», поскольку она имеет для говорящего статус субъективной истины.

[Разлогова 2005] для описания семантики вводных модальных слов предлагает иерархическую систему, в которой истинностные значения «Истина», «Ложь», «Неопределенность» распределяются по «логико-когнитивным состояниям» (они же «степени достоверности»): «Знание», «Уверенность», «Предположение».

4 Описание предлагаемой модели субъективной достоверности

4.1 Пропозиция как объект оценки

Пропозициональное содержание высказывания, или пропозиция, — это та часть смысла высказывания, которая «является естественным аргументом модальных операторов и предикатов пропозициональной установки» [Падучева 1985]. Сама пропозиция, выступающая объектом оценки СД, не содержится в предложении в явном виде.

Представление пропозиционального содержания

Чтобы сделать пропозицию объектом анализа, нужно ее каким-то образом «материализовать». Однозначного ответа на вопрос о том, как это сделать, не существует. Мы выбрали способ представления, который кажется довольно наглядным и при этом наиболее подходящим для рассуждений о СД: пропозиция представляется как значение придаточного предложения с союзом «то, что» (ср. у Куайна [Куайн 2000] пропозиция — это «абстрактный объект, мыслимый как то, что обозначается» ЧТО-придаточным (*that-clause*); о фактообразующем «то-что»-значении см. также [Арутюнова 1988]). Представленную в таком виде пропозицию можно будет естественным образом рассматривать как аргумент параметра, оценивающего ее с т. зр. СД.

Например:

- (7) *Невероятно, чтобы метеоритный дождь шел несколько тысячелетий* — пропозиция, которая будет оцениваться с т. зр. СД, — ‘то, что метеоритный дождь шел несколько тысячелетий’

Отрицание в пропозиции

Чтобы не смешивать на этапе оценки СД отрицательную полярность пропозиции (наличие в ней отрицания) и ее ложность, отрицание считаем элементом пропозициональной структуры [Падучева 1985].

- (8) *По-моему, твоя песня не понравилась Алле* — оценивается пропозиция ‘то, что твоя песня не понравилась Алле’

Модализованные пропозиции

Модальные слова считаем частью модализованной пропозиции. Исключение составляют операторы эпистемической возможности и необходимости, которые мы рассматриваем как эпистемические маркеры, ср. (9) и (10).

- (9) *Поэтому не считаю, что такие льготы должны быть одинаковыми во всех субъектах* — оценивается пропозиция ‘то, что такие льготы должны быть одинаковыми во всех субъектах’, *должны* — показатель деонтической необходимости.
- (10) *Это могло быть связано с длительностью процедур госзакупок* — оценивается пропозиция ‘то, что это было связано с длительностью процедур госзакупок’, *могло* — маркер эпистемической возможности.

Далее в примерах вместо восстановления оцениваемой пропозиции в явном виде ограничимся подчеркиванием содержащей ее части предложения.

Ограничение на способ выражения СД

В настоящей работе рассматриваются только те пропозиции, которые находятся в сфере действия лексически выраженных показателей СД.

Типы пропозиций

В интересах решаемой задачи релевантным представляется рассматривать пропозиции в плоскости «верифицируемость — неверифицируемость (оценочность)»:

- **Верифицируемые.** «Соотносятся с настоящим, прошлым или будущим положением дел в мире и в принципе могут быть верифицированы в момент произнесения соответствующего высказывания» [Дмитровская 1988]. Это прежде всего «событийные» пропозиции по [Шмелёва 1994] — они «портретируют» действительность;
- **Скорее верифицируемые.** «Логические» — «представляют результаты умственных операций и сообщают о некоторых установленных признаках, свойствах, отношениях» [Шмелёва 1994]: «анкетная» характеристика, идентификация, классификация (таксономическая идентификация), релятивные пропозиции;
- **Скорее неверифицируемые.** Рациональные (мотивированные) оценки — например, морально-этические;
- **Неверифицируемые.** Представляют свойства, оценку которых нельзя опровергнуть (хотя ее можно оспорить), — сенсорно-вкусовые, интеллектуальные, эмоциональные, эстетические (об оценочных суждениях см. [Дмитровская 1988]). Эти оценки не требуют мотивировки, поскольку «им не могут быть сопоставлены некоторые качества оцениваемого объекта [...] Оценка прямо проистекает из того ощущения, которое, независимо от воли и контроля, испытывает человек» [Арутюнова 1982].

4.2 Оцениваемый параметр — эпистемический статус

Для квалификации пропозиции с т. зр. СД мы ввели параметр *эпистемический статус* (ЭС). ЭС всегда привязан к субъекту. В настоящей работе, как уже говорилось, рассматриваются только те случаи, в которых ЭС отражает квалификацию пропозиции говорящим⁵.

Качественным ядром ЭС выступает истинностное значение («истина» или «ложь»), отнесенное к одной из трех эпистемических категорий (когнитивных установок): не-мнение («знание») и два вида мнения — «полагание/оценка» и «гипотеза» (Рисунок 1). «Истина» и «ложь» в каждой категории приобретают специфический смысл, который отражается в их условных названиях: в категории «знание» мы имеем дело с «”объективной”⁶ истиной» и «”объективной” ложью», в

⁵ В случае передачи чужой речи один и тот же маркер может выражать эпистемические установки говорящего и другого субъекта. В таких случаях мы сейчас ограничиваемся установкой говорящего (см. пример с *соврал* в таблице 2). Примеры типа «я (тогда) думал, что Р», «я соврал, что Р», в которых говорящий-в-настоящем отделяет себя от субъекта речемыслительного акта (себя же в прошлом), приравниваются к случаям передачи чужой речи.

⁶ «Объективный» взято в кавычки, поскольку означает здесь всего лишь «не имеющий признаков субъективного мнения».

категории «полагание/оценка» — с «субъективной истиной» и «субъективной ложью» и в категории «гипотеза» — с «гипотетической истиной» и «гипотетической ложью».



Рисунок 1: Когнитивные установки говорящего и соответствующие им истинностные значения

Количественные индексы

Гипотетические истина и ложь всегда имеют при себе индекс вероятности — одно из значений шкалы: «1», «весьма вероятно», «вероятно», «возможно».

Кроме того, истинностное значение в любой категории может быть охарактеризовано по признаку с обобщенной количественной семантикой (частные проявления которой — категоричность, уверенность, полнота, надежность) и значениями «сильн.», «слаб.».

4.3 Значения эпистемического статуса и принципы их присваивания

Категория «знание»: «“объективная” истина», «“объективная” ложь»

Главное, что отличает значения ЭС в этой категории, — отсутствие признаков собственного мнения говорящего. Информацию о положении дел, обозначенном пропозицией Р, говорящий получил, если можно так выразиться, «в готовом виде» — посредством чувственного восприятия (зрение, слух, обоняние), из собственного опыта, памяти, от других лиц или из сторонних источников и т. п. (примеры в таблице 2). Сама пропозиция обычно верифицируема, но не обязательно (ср., напр., *Я слышал, что это вкусно*). ЭС «сильная истина» или «сильная ложь» в категории «знание» пропозиция получает, если говорящий снабдил ее явным указанием на соответствие или несоответствие действительному положению дел. ЭС «слабая истина» или «слабая ложь» — если говорящий допускает, что его знания неполны или не совсем надежны (ошибка восприятия, несовершенство памяти и т. п.). Что касается нейтральных «объективных» истины и лжи, то они обычно выражаются имплицитно, однако их можно усмотреть и в некоторых лексически маркированных случаях.

«объективная» ИСТИНА Г знает (думает, что знает), что Р истинно	количеств. признак (полнота / категоричность)	«объективная» ЛОЖЬ Г знает (думает, что знает), что Р ложно
<i>Не отрицаю, бывают моменты, когда фонари на улицах не горят из-за поломок.</i>	нейтральная	<i>Флавио не моргнув глазом соврал⁷, что АЛОНСО самостоятельно провернул переговоры с МАКЛАРЕНОМ</i>
<i>Во-вторых, насколько я понял, люди, у которых врач заподозрил наличие зависимости, должны будут сдать анализ за свой счет и пройти тест.</i>	слабая	<i>Не слышал, чтобы у перевозчиков возникали сложности с выплатой налогов.</i>
<i>Когда окончательно подтвердилось, что выживших нет, это был не просто день официального национального траура – это было горе всей страны.</i>	сильная	<i>Неправда, что русские генералы мечтали о броске к Атлантике, Средиземноморью, Ла-Маншу.</i>

Таблица 2: Примеры «объективных» ЭС пропозиции

Категория «полагание/оценка»: «субъективная истина», «субъективная ложь»

Пропозиция Р соотносится не с реальной действительностью, а с точкой зрения говорящего, с его собственной квалификацией некоторой ситуации или объекта. При этом говорящий не строит вероятностные догадки об истинности или ложности Р (как было бы в случае гипотезы, см. далее), а высказывает свое мнение об этом с большей или меньшей уверенностью, категоричностью. Пропозиция, имеющая статус «субъективная истина» или «субъективная ложь», не верифицируема, ее нельзя просто подтвердить или опровергнуть. Но с ней можно согласиться или не согласиться. Примеры — в таблице 3.

⁷ Говорящий знает, что это не так.

субъективная ИСТИНА Г считает (чувствует), что Р истинно	количеств. признак (уверенность / категоричность)	субъективная ЛОЖЬ Г считает (чувствует), что Р ложно
<i>Полагаю, что все ФАПы должны быть в прямой коммуникации и системе консультаций с краевыми</i>	нейтральная	<i>Я не согласен с тем, что Дании уже пора открывать свои границы со Швецией, так как ситуация с коронавирусом в Швеции все еще не находится под контролем</i>
<i>Мне кажется, <u>ничего страшного не происходит</u></i>	слабая	<i>Не думаю, что <u>это нарушение прав человека.</u></i>
<i>Я уверен, что <u>с предателем надо поступать, как с изменником Родины.</u></i>	сильная	<i>"Я решительно возражаю против того, что <u>это было соразмерное применение силы в той ситуации</u>", - сказал Аррадондо.</i>

Таблица 3: Примеры субъективных ЭС пропозиции

Категория «гипотеза»: «гипотетическая истина», «гипотетическая ложь»

Говорящему не известно истинное положение дел, обозначенное пропозицией Р (или не может быть известно — судить об этом нам позволяет дискурсивный контекст, в т. ч. здравый смысл и другие фоновые знания). Говорящий может лишь высказать предположение об истинности или ложности Р, основанное на логическом выводе из имеющихся в его распоряжении данных или на интуитивной догадке (ср. определение гипотезы в [Булыгина, Шмелёв 1997]). Истинность или ложность Р в этом случае поддается оценке в терминах вероятности. Для количественной оценки вероятности мы используем четырехзначную шкалу: «1», «весьма вероятно», «вероятно», «возможно». Значение «возможно» интерпретируется не как минимальное значение вероятности, а как неопределенное ее значение, отличное от нуля. Исходим из того, что эпистемическая возможность предполагает ненулевую вероятность, а эпистемическая невозможность — нулевую. Для выражения невозможности (нулевой вероятности) «истины» используем «ложь» с вероятностью «1» (таблица 4).

Гипотетическая ИСТИНА	Вероятность	Гипотетическая ЛОЖЬ
<i>Закручивать гайки для легальных интернет-видеосервисов бессмысленно, потому что <u>нарушения происходят явно не на их платформах.</u></i>	«1»	<i>Не может такого быть, чтобы <u>руководство не знало о существовании такой схемы.</u></i>
<i>Простите мне пессимизм, но я практически уверен, что <u>судьба московского особняка купца Булошников на Большой Никитской, 17, решена</u> и никакие слушания (назначены на четверг, 17 января) ее уже не изменят.</i>	«весьма вероятно»	<i>Очень сомнительно, что <u>сам момент разрушения фигуры никто не видел, а значит, горожане не остановили порчу декоративной фигуры.</u></i>
<i>Судя по всему, <u>переговорный процесс по конкретным проблемам продолжается.</u></i> <i>В него выстрелил Соколов, <u>настоящее имя которого, предположительно, Вадим Красиков.</u></i> <i>Наверное, <u>есть какие-то инвесторы, предприниматели, которые именно такие: построили – продали.</u></i>	«вероятно»	<i>Маловероятно, что <u>люди способны выработать антитела, которые будут гарантировать их иммунитет к короновирусу в течение долгого времени.</u></i> <i>Я сомневаюсь, что <u>у мошенников есть доступ к этим базам.</u></i>
<i>Нельзя исключать, что <u>некая общественная организация собирает информацию для подготовки соответствующей законодательной инициативы.</u></i> <i>Однако существует вероятность, что на самом деле найдены останки Мариш дель Риччо — второй обительницы монастыря, которая, как и Лиза Герардини, не была монахиней, но также удостоилась особого захоронения из-за своего благородного происхождения.</i>	«возможно»	<i>Не уверен, что <u>даже налоговые органы в полной мере осведомлены об их доходах.</u></i> <i>Нельзя с уверенностью утверждать, что <u>Эци страдал при жизни от гастрита, ведь слизистая оболочка желудка не сохранилась.</u></i>

Таблица 4: Примеры гипотетических ЭС пропозиции

Примечание. Для случаев, когда говорящий в той или иной форме отказывается брать на себя эпистемическую ответственность за пропозицию, предусмотрено особое значение ЭС — «неизвестно», например:

- (11) ***Не берусь утверждать, что Пригожин причастен и к указанному сайту в сети «ВКонтакте».***

5 Разметка датасета

Пригодность предложенной модели СД оценивалась путем разметки текстового материала, который представляет рассматриваемое явление в его многообразии, с последующим вычислением уровня согласия аннотаторов.

5.1 Принципы разметки субъективной достоверности

Разметке подлежат только лексически маркированные случаи выражения СД, удовлетворяющие следующим условиям:

1. **Эксплицитность.** СД оцениваемой пропозиции должна быть выражена явными лексическими средствами — словами и конструкциями, которые содержат в своем значении эпистемический компонент.⁸ (В [Зализняк 1992] эпистемический компонент определяется как 'X знает, что P', 'X считает, что P', 'X считает, что P вероятно', 'X считает, что P возможно').

2. **Перволичность.** Субъектом эпистемической установки, выраженной маркером, является говорящий. Установки других субъектов — в случаях передачи чужой речи — сейчас не рассматриваются.

5.2 Датасет

С помощью описанного метода был размечен датасет из 510 текстовых фрагментов объемом одно-два предложения. В качестве источника использованы общий и газетные подкорпуса НКРЯ (<https://ruscorpora.ru/>).

Основной целью при отборе примеров был максимальный охват способов явного выражения СД. (Поскольку количественное распределение показателей в реальных текстах никак не учитывалось, о репрезентативности датасета мы говорить не можем.) Особое внимание уделялось потенциально неоднозначным показателям. Для каждого показателя (представителя группы синонимичных показателей) по возможности отбирались примеры, содержащие разные типы пропозиций.

Эпистемические показатели в плане выражения были разнообразны:

- предикаты внутреннего состояния, у которых эпистемический компонент является частью толкования (*я считаю..., думаю, что..., сомневаюсь, что...*) и синонимичные им конструкции (*есть сильные сомнения, что...*);
- предикаты восприятия (*я не видел, чтобы...*);
- модальные глаголы со значением эпистемической возможности и необходимости (*может, должен*);
- вводные слова (*возможно, по-моему, кажется*);
- наречия со значением оценки вероятности (*наверняка, маловероятно*);
- частицы (*едва ли, якобы*);
- эксплицитное указание на истинность или ложность ассоциированной пропозиции (*это правда, что..., информация о... фейк*), а также частичный или полный отказ говорящего от эпистемической ответственности (*не берусь утверждать...*);
- указание на источник информации или на отсутствие такового (*доказано, что..., нет данных о том, что...*)

и др.

Разметка производилась в таблице, в столбцах которой аннотаторы для каждого примера записывали значения атрибутов ЭС (истинностное значение, установка, количественные индексы).

Датасет размещен в открытом доступе: <https://github.com/VozdvIN/AIReC-EpistemicDataSet/releases>

5.3 Оценка уровня согласия аннотаторов

Разметка датасета выполнялась тремя аннотаторами, имеющими лингвистическое образование. Для них была подготовлена подробная инструкция с содержательным описанием значений СД и показательными примерами.

Для оценки уровня согласия между аннотаторами был использован показатель Каппа Флейса [Fleiss 1971], далее KF. KF выражает степень отличия фактического распределения мнений аннотаторов относительно полностью случайного выбора значений.

Этот показатель рассчитывался дважды: для сокращенной и для полной версии разметки.

⁸ Вопросы имплицитного выражения СД требуют отдельного рассмотрения.

При расчете KF для сокращенной разметки оценивалось согласие для пар «истинностное значение, категория» (например, «истина-знание», «ложь-полагание» и др., а также внекатегориальный статус «неизвестно»). «Количественные» признаки не учитывались. В этом случае KF=0,93.

При расчете KF для полной разметки оценивалось согласие по значению ЭС целиком, с учетом «количественных» показателей. Например, «ложь-полагание (нейтральная)» и «ложь-полагание слабая» в этом случае считались за несовпадающие значения. Для этого варианта KF=0,86.

В обоих случаях значение KF интерпретируется как «очень высокая степень согласия» [Fleiss 1981].

Типичные случаи расхождения аннотаторов в выборе категории имеют место, когда категориальная неоднозначность показателя соединяется с неочевидным характером пропозиции. Пример сложного выбора между субъективным и гипотетическим статусом:

- (12) *Я убежден в том, что у России достаточно средств для того, чтобы не уступить в конкуренции.*

Пример размечен как «сильная истина-полагание/оценка» и «гипотетическая истина с вероятностью 1». Неясно, то ли автор-говорящий знаком с реальным положением дел (в данном случае — знает, какими средствами располагает Россия) и дает этому положению дел свою субъективную оценку (этих средств достаточно, чтобы...). То ли положение дел автору не известно, но он предполагает что оно таково, что он бы его оценил как достаточное для того, чтобы...). Вторая интерпретация сродни тому, что [Дмитровская 1988] называет оценкой-предположением («Я думаю, что книга интересная»).

Аналогичный пример — размечен как «гипотетическая ложь (вероятна)» и как «слабая ложь-полагание/оценка»:

- (13) *Мне не кажется, что это станет серьезным подспорьем в пополнении бюджета* [Заголовки публикации: «Эксперт оценил пользу для бюджета от легализации продажи «красивых» номеров»]

Пока не решено, как лучше поступать в случаях, когда однозначно установить категорию значения не удастся. Однако мы в любом случае исходим из того, что модель языкового явления должна давать возможность различать то, что в языке регулярно различается.

6 Заключение

В статье представлена модель субъективной достоверности, в которой для различения случаев маркированной достоверности предлагается выявлять когнитивную установку говорящего.

Модель была испытана при аннотировании датасета из 510 примеров тремя аннотаторами. Несмотря на высокие показатели согласия, в некоторых случаях решение давалось аннотаторам нелегко. Спорные случаи обсуждались, и по результатам обсуждения разметка могла быть частично исправлена.

Задачи на ближайшее будущее:

- систематизировать сложные и спорные случаи;
- рассмотреть возможность дифференциации значений внутри когнитивных установок;
- существенно увеличить объем размеченного материала;
- задействовать большее число аннотаторов.

References

- [1] Antici F. et al. A corpus for sentence-level subjectivity detection on english news articles // Proceedings of the LREC-COLING 2024. P. 273–285. <https://doi.org/10.48550/arXiv.2305.18034>
- [2] Diab M. et al. Committed belief annotation and tagging // Proceedings of the Third Linguistic Annotation Workshop (LAW III). 2009. — P. 68–73. <https://aclanthology.org/W09-3012.pdf>
- [3] Fleiss J. L. Measuring nominal scale agreement among many raters // Psychological Bulletin 76 (5). 1971. — P. 378–382.
- [4] Fleiss J. L. Statistical methods for rates and proportions. New York: John Wiley & Sons, 1985. — P. 38–46.
- [5] Glòria Vázquez García, Ana María Fernández Montraveta. Annotating factuality in the TAGFACT corpus // Multiperspectives in analysis and corpus design. — 2020. P. 115–127.
- [6] Ankita Gupta, Su Lin Blodgett, Justin H Gross, Brendan O’Connor. Examining Political Rhetoric with Epistemic Stance Detection // Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS), November 7, 2022. — Pp. 89–104.
- [7] Lee K. et al. Event detection and factuality assessment with non-expert supervision // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. — 2015. — P. 1643–1648. <https://yoavartzi.com/pub/lacz-emnlp.2015.pdf>
- [8] Chunyang Li, Hao Peng, Xiaozhi Wang, Yunjia Qi, Lei Hou, Bin Xu, and Juanzi Li. 2024. MAVEN-FACT: A large-scale event factuality detection dataset // Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA. — Association for Computational Linguistics. — P. 11140–11158.
- [9] Salvador Lima, Naiara Perez, Montse Cuadros, German Rigau. NUBES: A Corpus of Negation and Uncertainty in Spanish Clinical Texts // Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020). — Marseille, 11–16 May 2020. — P. 5772–5781. <https://aclanthology.org/2020.lrec-1.708.pdf>
- [10] Minard A. L. et al. MEANTIME, the NewsReader multilingual event and time corpus // Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16). — 2016. — P. 4417–4422. http://www.lrec-conf.org/proceedings/lrec2016/pdf/488_Paper.pdf
- [11] Marco Rovera, Serena Cristoforetti, Sara Tonelli. ModaFact: Multi-paradigm Evaluation for Joint Event Modality and Factuality Detection // Proceedings of the 31st International Conference on Computational Linguistics. — January 19–24, 2025. — P. 6378–6396.
- [12] Saurí, R., Pustejovsky, J. FactBank: a corpus annotated with event factuality // Lang Resources & Evaluation 43, 227–268 (2009). <https://doi.org/10.1007/s10579-009-9089-9>
- [13] Saurí R., Pustejovsky J. Are you sure that this happened? Assessing the factuality degree of events in text // Computational linguistics. — 2012. — 38(2) — P. 261–299.
- [14] Арутюнова Н.Д. Аксиология в механизмах жизни и языка // Проблемы структурной лингвистики. — М.: Наука, 1984.
- [15] Арутюнова, Нина Давидовна. Типы языковых значений [Текст] / Н. Д. Арутюнова; Оценка, событие, факт ; Отв. ред. Г. В. Степанов ; АН СССР, Ин-т языкознания. — М.: Наука, 1988.
- [16] Булыгина Т. В., Шмелёв А. Д. Языковая концептуализация мира (на материале русской грамматики). — М.: Школа «Языки русской культуры», 1997. — 576 с.
- [17] Дмитровская М. А. Знание и мнение: образ мира, образ человека // Логический анализ языка. Знание и мнение. — М.: Наука, 1988. — С. 6–18.
- [18] Зализняк Анна А. Исследования по семантике предикатов внутреннего состояния. — Verlag Otto Sagner, München, 1992.
- [19] Куайн У. В. О. Слово и объект. — М.: Логос, Праксис, 2000.
- [20] Падучева Е.В. Высказывание и его соотносительность с действительностью (Референциальные аспекты семантики местоимений) — М.: Наука, 1985.
- [21] Разлогова Е. Э. Логико-когнитивные и стилистические аспекты семантики модальных слов: автореф. дис. ... д-р филол. наук. — Москва, 2005.
- [22] Шмелёва Т.В. Смысловая организация предложения и проблема модальности // Актуальные проблемы русского синтаксиса / Под ред. К.В. Горшковой, Е.В. Клобуковой. — М., 1984. — С. 78–100.
- [23] Шмелёва Т. В. Семантический синтаксис: текст лекций из курса «Современный русский язык» / Т. В. Шмелёва. — Красноярск: Красноярский государственный университет, 1994. — 46 с.
- [24] Яковлева Е. С. Фрагменты русской языковой картины мира (модели пространства, времени и восприятия) / Е. С. Яковлева. — Москва : Гнозис, 1994. — 344 с. (Язык. Семиотика. Культура).