FactBank 1.0
Annotation Guidelines

Roser Saurí
Brandeis University

May 18, 2008

Contents

1 Introduction

This document provides the guidelines for annotating information related to the factuality of events. Section 2 introduces the basic notions that are of relevance here (*event factuality, factuality markers* and *sources*), and section 3 covers the specifics of each annotation task.

2 Relevant Notions

2.1 Event Factuality

Eventualities in discourse are characterized with different degrees of factuality. In some contexts, the factual status of events is presented with absolute certainty. Events are then depicted as *facts* –that is, as situations that happened or took place in the world (1a)– or *counterfacts* –situations that did not take place in the world (1b).

(1) a. Five other U.N. inspection teams <u>visited</u> a total of nine other sites.

  b. The size of the contingent was not <u>disclosed</u>.[1]

 Other contexts introduce different shades of uncertainty. If the contextual polarity is positive, events are then qualified as being *possibly factual* (2a); by contrast, in contexts of negative polarity events are presented as being *possibly counterfactual* (2b).

(2) a. United States may <u>extend</u> its naval quarantine to Jordan's Red Sea port of Aqaba.

  b. They may not have <u>enthused</u> him for their particular brand of political idealism.

 We will therefore understand **event factuality** as the level of information expressing whether events in discourse refer to real situations in the world (facts), have no real counterpart (counterfacts), or are of an uncertain nature (possibilities).

2.2 Factuality markers

Event factuality is expressed in text by means of what we call **factuality markers**. There are a number of them. For example, a negative polarity particle expresses the counterfactive nature of an event (1b), whereas a modal auxiliary such as *may* introduces uncertainty (2).

 Predicates subcategorizing for a clause of some sort are another type of marker. In (3), for instance, the verbal predicate *said* (e_0) has an effect on the factual status of its embedded event (e_1). Specifically, e_0 frames e_1 within an evidential context, and as a result, e_1 is presented as a fact according to the reporting source, Slobodan Milosevic's son.

(3) Slobodan Milosevic's son **<u>said</u>**$_{e_0}$ Tuesday that the former Yugoslav president had been <u>murdered</u>$_{e_1}$ at the detention center of the U.N. war crimes tribunal in The Hague. [Herald Tribune, 03/14/2006]

 Sentence (4) provides another example of predicates functioning as factuality markers. Here, the predicate *attempt* qualifies the climbing event with a degree of uncertainty.

(4) George Mallory and Andrew Irvine first **<u>attempted</u>**$_{e_0}$ to <u>climb</u>$_{e_1}$ Everest in 1924. [EverestNews.com, 2004]

---

[1]Here and throughout this document, events in the examples will be identified by marking only the verbal, nominal, or adjectival predicates expressing them, along the lines of TimeML annotation. Furthermore, only the events relevant at each point of the discussion will be identified for clarity. The examples have been extracted from the British National Corpus, the American National Corpus, TimeBank, and Google.

2.3 Sources

The factuality value assigned to events in text is always relative to one or more participants who commit to that value at a particular point in time. They are referred to as **sources**.

By default, events mentioned in discourse, be it oral or written, have always an implicit source corresponding to the author of the text (from here on referred to as s_0). In (5), for example, CNN commits to the factuality of Milosevic being on trial.

(5) Milosevic was on $\underline{trial}_{e_0}$ for 66 counts of genocide, crimes against humanity and war crimes in the Balkans during the 1990s. [$\mathbf{CNN}_{s_0}$, 03/12/2006]

Additional sources can be introduced by means of what we call *source-introducing predicates* (SIPs). These are, for instance, predicates referring to reporting events (such as *say, tell*), but also predicates of knowledge and opinion (e.g., *believe, know*), psychological reaction (*regret*), etc.

The new source tends to be expressed as the grammatical subject of the SIP, and play a role in assessing the factuality of the event the SIP selects for. Let's consider example (3) in detail, repeated below as (6).

(6) **Slobodan Milosevic's son**$_{s_1}$ $\underline{said}_{e_0}$ Tuesday that the former Yugoslav president had been $\underline{murdered}_{e_1}$ at the detention center of the U.N. war crimes tribunal in The Hague. [**Herald Tribune**$_{s_0}$, 03/14/2006]

It contains two event expressions: e_0 and e_1. We need to find the relevant sources for each of them. By default, the main event of the sentence, e_0, has the text author, s_0, as its relevant source –that is, as the source expressing either commitment or non-commitment towards its factual status. Now, what about e_1? It it is selected by e_0, in fact a SIP introducing a new source in discouse: Slobodan Milosevic's son (s_1). Hence, event e_1 has two sources: the text author (s_0) and Slobodan Milosevic's son (s_1). Observe that Milosevic's son commits to e_1 as a fact, whereas the author keeps a neutral attitude towards that same event.[2]

Strictly speaking, however, we do not have direct access to the factual assessment of e_1 by Milosevic's son. We know about Milosevic's son's perspective only according to what s_0, the author, asserts. Hence, we need to appeal to the notion of *nested source*. That is, the second source in (6) is not Milosevic's son, but Milosevic's son (s_1) according to the author (s_0), which we will represent as: $s_{1\_}s_0$. To sum it up, the relevant sources for each of the events in (6) are:

(7) a. e_0: *said* $< s_0 >$
    b. e_1: *had been murdered* $< s_0 >, < s_{1\_}s_0 >$

3 Annotation Tasks

3.1 Task 1: Identification of Source-Introducing Predicates

Given a text with the events already recognized, identify the set of predicates that satisfy the following criteria:

---

[2] For readers, the factuality status of e_1 depends to a greater or lesser degree on the reliability of the reporting source. Followers of the ex-president of Serbia may take the statement of Milosevic's son as expressing a fact, whereas for other people that may be utterly false. In the present work, we are not interested in assessing the factuality of events given our own beliefs and knowledge, but based on the information provided by the text.

1. EVENT STATUS: They are recognized as **events**.

2. PART OF SPEECH: They are expressed by either a **verb** (*think, say*), a **noun** (*report, doubt*), or an **adjective** (*aware, confident*).

3. LEXICAL SEMANTICS: As part of their meaning, **they express the stance or commitment of a given source towards the factuality of the embedded event**. For example, *suspects* in (8a) presents Danielle as committing to the factual status of event e_1 (Betina leaving) as being lower than the factual status of the same event in (8b), where it is embedded under *knows*. In other words, in (8a) Danielle considers event e_1 as being just a possibility, whereas in (8a) she contemplates it as a fact.

(8) a. Danielle $\underline{\textbf{suspects}}_{e_0}$ that Betina $\underline{\text{left}}_{e_1}$ the country in June.

  b. Danielle $\underline{\textbf{knows}}_{e_0}$ that Betina $\underline{\text{left}}_e$ the country in June.

Our predicates can be classified as belonging to one of the following classes:

Predicates of report. E.g., *say, tell, add; claim, argue* –even if they express report by other means than oral; e.g., *write, publish, post.*

Predicates of knowledge. They can express the state of having knowledge (*know, understand, remember*), acquiring knowledge (*learn, find out, discover*), loosing knowledge (*forget*), admitting knowledge (*acknowledge, accept, admit*).

Predicates of belief and opinion. E.g., *think, find, consider, guess, predict, estimate, suggest.*

Predicates of doubt. E.g., *doubt, wonder, ask* (they generally subcategorize for a *wh*-clause complement).

Predicates of perception. E.g., *see, hear, feel.*

Predicates expressing proof. E.g., *prove, show, support, explain.*

Predicates expressing some kind of inferencing process. E.g., *infer, deduce; conclude, decide (that); mean, suggest* (as in: *it means/suggests that*); *appear, seem* (as in: *it seems/appears that*).

Predicates expressing some psychological reaction as a result of an event or situation taking place: *regret, be glad/pleased (that), like (that), love (that).*

Other classes are also possible, as long as the predicate satisfies the other criteria.

DEALING WITH POLYSEMY: Many of these predicates are polysemous. In some cases, the different senses all fall in the classes listed above. For example, *explain* denotes both a reporting act (9a) and a proving state (9b).

(9) a. She explained that she was going to stay with her sister.

  b. This could explain why educational interventions haven't been more successful.

In other cases, some of the senses may not be classifiable under any of the classes above. *Promise* and *agree* can be used as reporting predicates:

(10) a. Clinton <u>promised</u> that he had totally changed. ($\approx$Clinton said that...)

    b. CBS <u>agrees</u> that online video is not the enemy. ($\approx$CBS says it is true that...)

Yet, they can also be used to express the attitude of the subject towards a potential future event (11). In that case, the complement is generally realized by means of an infinitival clause –although a *that*-clause is also possible.

(11) a. Prime Minister John Howard , a monarchist himself, <u>promised</u> to put the question to a national referendum next year.

    b. Iraq <u>agreed</u> to give inspectors full access to eight of Saddam Hussein's presidential palaces.

In these second senses, *agree* and *promise* are similar to other predicates that subcategorize for *to*-clauses as well (such as *want, force, offer*, or *need*), which are not SIPs. They all share the following features:

1. The complement event is always a future possibility relative to the embedding event.

2. The attitude of the subject towards the complement event is different than the attitude in the predicates classified above (predicates of report, belief, knowledge, etc.). In those predicates, the subject attitude is epistemic in nature; that is, concerning the degree of certainty that an event has taken (or will take) place –or, in our words, concerning the factual status of events.

    By contrast, predicates like *want, force*, and *offer* present their subjects as:

- Either having some degree of responsability on the potential event (e.g., *agree/ promise/offer to go*; *force somebody to go*), or:
- Being in a greater or lesser favorable state towards the accomplishment of the embedded event (*need/want to go*).

Neither these predicates (*want, force, offer, need*), nor the senses of *promise* and *agree* in examples (11), among other predicates, will be identified as SIPs.

4. COMPLEMENT: They take a **complement which expresses another event**. Complement of **verbal SIPs** can be of any of the categories listed below (where the complement is marked with square brackets, its main event-denoting expression identified as s_1, and the SIP as s_0):

    A *that*-clause. E.g., *He <u>thinks</u>$_{e_0}$ [several steelmakers will <u>report</u>$_{e_1}$ actual losses through the third quarter of 1990].*

    An infinitival clause: *Gore never <u>claimed</u>$_{e_0}$ [to have <u>invented</u>$_{e_1}$ the Internet].*

    An *ing*-clause: *He was caught when a flight attendant <u>saw</u>$_{e_0}$ him [<u>attempting</u>$_{e_1}$ to light a match on the tongue of his shoe].*

    An NP headed by an event-denoting noun. E.g., *Coast Savings Financial Inc. <u>reported</u>$_{e_0}$ [a third-quarter <u>loss</u>$_{e_1}$].*

    Possibly other constructions, as long as their head expresses an event.

    In some cases, the event complement is not used as a direct complement of the SIP, but as a predicative one. Contrast examples in (12) versus (13):

(12) a. He **thinks**$_{e_0}$ [several steelmakers will report$_{e_1}$ actual losses through the third quarter of 1990].

b. You made the **comment**$_{e_0}$ [that the Gulf War coalition in '91 was$_{e_1}$ far stronger than this].

c. Washington was **aware**$_{e_0}$ [that a deadly tidal wave was building$_{e_1}$ up in the Indian Ocean].

(13) a. What he **thinks**$_{e_0}$ is [that the Vail Valley and Colorado are losing$_{e_1}$ the Western heritage upon which the West was founded].

b. The **news**$_{e_0}$ about real estate here was [that the sky was$_{e_1}$ the limit the highest prices in the world].

c. What he is **confident**$_{e_0}$ about is [that the opposition will be united$_{e_1}$ in the end].

It is also possible that the SIP complement is expressed by a pronoun referring to a previously mentioned event. For instance, in the next example, *this* refers to e_1.

(14) They believed it will be$_{e_1}$ always up forever. Nobody believes$_{e_0}$ **this** any more.

5. COGNITIVE AGENT: In addition to that complement, they also select for an **argument expressing a cognitive agent** (or cognizer). Part of the predicate semantics is expressing the stance of that cognitive agent towards the factual value of the complement event (recall the examples in (8)). Using our terminology, we say that the cognitive agent is introduced as a source of the factual status of the complement event.

In case of doubt, the presence of a cognitive agent can help to decide whether a given element is truly a SIP, precisely because SIPs are defined as *source-introducing* predicates. Note that the cognizer (or source) may correspond to a new discourse participant. In (15a), for example, *Milosevic's son* is introduced as a source assessing the factuality of e_1, in addition to the text author. But in other cases, the cognizer corresponds to a previously known cognizer, such as in (15b), where the pronoun *I* refers to the text author.

(15) a. **Slobodan Milosevic's son** said$_{e_0}$ Tuesday that the former Yugoslav president had been murdered$_{e_1}$.

b. **I** think$_{e_0}$ it's not going to change$_{e_1}$ for a couple of years.

Here are some guidelines for identifying cognizers:

1. In most of the cases, cognitive agents are realized as the **grammatical subject** of the predicate. In the example below, where *denied* is the SIP, the individuals referred to by the grammatical subject, *Colin Powell* and *Condoleezza Rice*, are both presented as sources of e_1.

   (16) In mid-2001, **Colin Powell** and **Condoleezza Rice** both publically denied$_{e_0}$ that Iraq had$_{e_1}$ weapons of mass destruction.

2. If the SIP candidate is embedded under another predicate, the cognitive agent may correspond to one of the arguments of that outmost embedding predicate (generally, its subject). For example, in (17) *Lady Charlotte Wynn* is the cognitive agent introduced by the SIP *regretted*, but also by the second SIP *learning*, which is embedded under the former.

   (17) **Lady Charlotte Wynn** regretted$_{e_0}$ learning$_{e_1}$ that Dundas and his relatives had received$_{e_2}$ upwards of £50,000 annually out of the public funds.

3. Sometimes, the cognitive agent is also expressed by means of an **oblique, possibly optional, complement**. In the following examples, the SIP is identified as e_0, the complement event as e_1, and the cogntive agent in bold face.

   (18) It $\underline{seems}_{e_0}$ to **him** that a girl's story about her goat and its butting $\underline{was}_{e_1}$ more important.

   (19) He was $\underline{told}_{e_0}$ by **Cheney** that Bush had $\underline{approved}_{e_1}$ a plan in which Libby would brief a specific New York Times reporter.

   Particular attention deserves the construction *according to NP*, where *according* is also a SIP:

   (20) $\underline{According}_{e_0}$ to **Jordanian officials**, a smaller line into Jordan $\underline{remained}_{e_1}$ operating.

4. Cognitive agents may also be introduced by nominal SIPs:

   (21) **Unisys Corp.**'s $\underline{announcement}_{e_0}$ Friday of a \$648.2 million $\underline{loss}_{e_1}$ for the third quarter showed that the company is moving even faster than expected.

5. In some constructions, the cognitive agents are not expressed in the sentence. Different situations require different treatments:

   Constructions with impersonal interpretation, where a cognizer different than the text author is implied. The SIP does not select for a cognizer as one of its arguments. However, there is an implicit reference to this participant.

   This is for instance the case of SIPs in past participle form. The cognitive agent is optionally expressed by means of a *by*-PP, but when not present in the sentence, a reference to an implicit cognizer is still understood. In the following example, for instance, the belief about Iraq's ability (e_0) is implicitly assigned to a specific group –probably experts.

   (22) Iraq is $\underline{believed}_{e_0}$ to $\underline{have}_{e_1}$ the ability to deliver chemical weapons in artillery shells.

   Nominal SIPs also tend to offer impersonal interpretations of this type:

   (23) There have been $\underline{reports}_{e_0}$ of wholesale $\underline{looting}_{e_1}$.

   Because the predicate is presuming an additional source, these cases WILL be identified as a SIPs (as long as all the remaining requirements apply).

   Constructions with impersonal interpretations, where no cognizer is implied. This is generally the case for certain predicates that optionally introduce a cognitive agent by means of an oblique complement. Contrast (24) with (25):

   (24) a. The move $\underline{seemed}_{e_0}$ $\underline{aimed}_{e_1}$ at heading off more trouble with Iran.
        b. Bush $\underline{seemed}_{e_0}$ to $\underline{suggest}_{e_1}$ that American forces could be in the gulf region for some time.

   (25) a. To **him**$_{s_1}$, the move $\underline{seemed}_{e_0}$ $\underline{aimed}_{e_1}$ at heading off more trouble with Iran.

b. To **some analysts**$_{s_1}$, Bush $\underline{\text{seemed}}_{e_0}$ to $\underline{\text{suggest}}_{e_1}$ that American forces could be in the gulf region for some time.

In (24), the SIP candidates, *seemed*, do not introduce any cognizer argument. Note that the grammatical subject of *seemed* in (24b) does not satisfy this role either, even though it expresses a cognitive individual. Compare it with example (25b): whereas here *seemed* expresses the factual status that *some analysts* assign to event e_1, in (24b) *seemed* does not denote the way *Bush* evaluates the event denoted by *suggest*.

If the phrase denoting the cognizer is not explicit in the text, the predicate will NOT be identified as a SIP.

Final Remarks   Other classes are possible, as long as the predicate expresses the assessment of a given cognizer with regards to the factual nature of a complement event.

3.2   Task 2: Identifying Sources

For this task, the text to annotate has the following elements already identified:

- The source-triggering predicate (SIP), colored in blue in figure 1.

- A set of potential candidates to be the source (or *cognitive agent*) introduced by the SIP. Their head will be colored in red and identified by a subscripted id.

The goal of this task is to select, from the set of source candidates colored in red, the source introduced by the blue-colored SIP. We will use the criteria 1-5 (pages 7-9) presented in Task 1 for deciding whether a predicate has a cognitive agent.

The annotation tool is presented in figure 1. The candidate is selected by clicking on the appropriate button at the right hand side of each sentence. Different situations may apply:

I. The source corresponds to one of the entities identified in red in the text.   Select the button corresponding to the candidate id. Refer to example in line #3 (s6) in figure 1. Some sentences may present additional complications:

- *The sentence presents two source candidates that semantically refer to the same entity.* Select the source grammatically introduced by the SIP; that is, the source that is expressed by either one of the SIP arguments or adjuncts.[3]

(26) So when Wong **Kwan**$_{c3}$ spent seventy million dollars for this house, **he**$_{c12}$ $\underline{\text{thought}}$ **it**$_{c14}$ was a great deal.

In (26), *Kwan* and *he* co-refer. We consider however that the source introduced by the SIP *thought* is *he*, its grammatical subject.

Because of this grammatically oriented apporach, relative pronouns are perfectly accepted as sources as well. Consider:

---

[3]In these and the following examples, source candidates will be in bold, and the SIPs underlined.
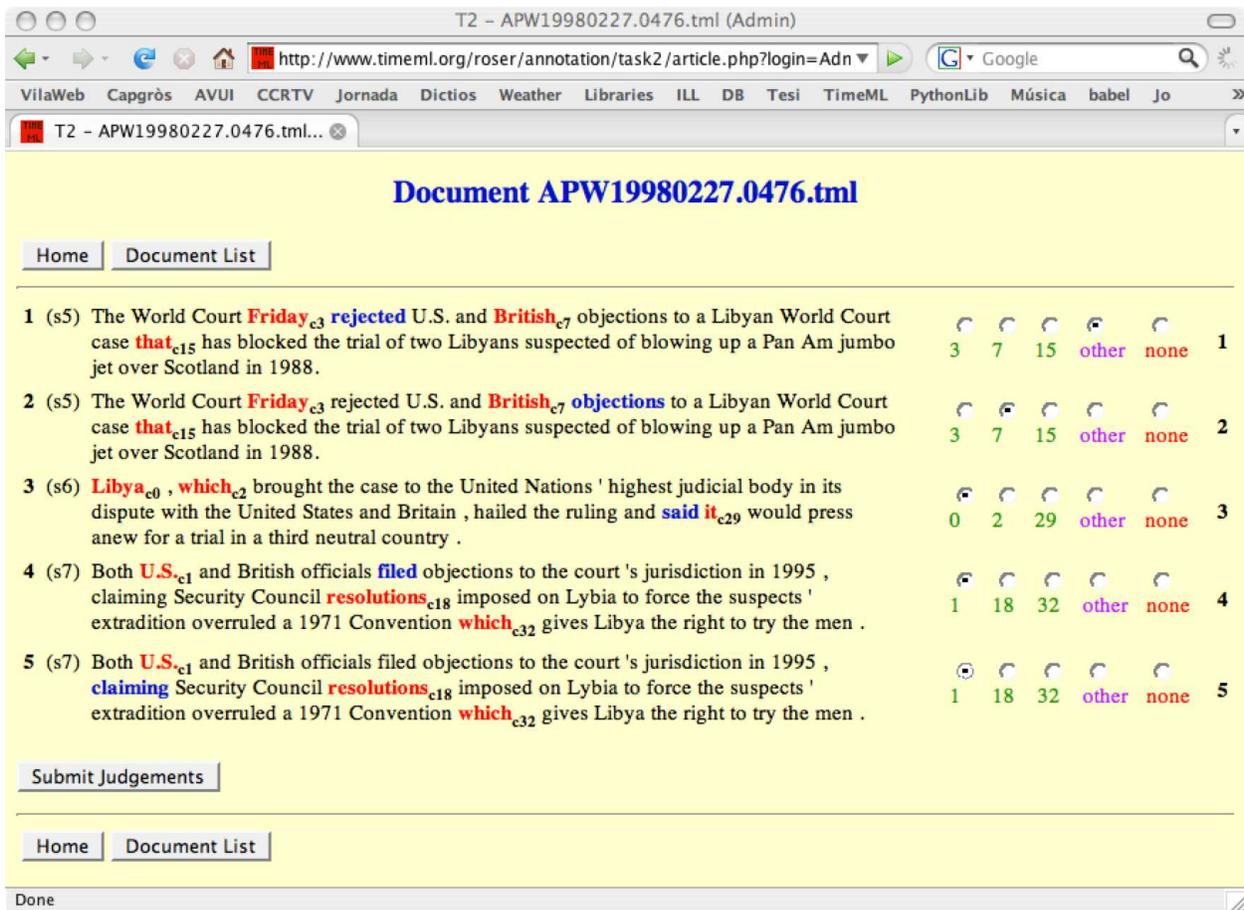
Figure 1: Task 2 annotation screen

(27) There was no independent confirmation of the report by the government-run news agency, **which**$_{c14}$ did not <u>say</u> when the reported **attempt**$_{c21}$ occurred.

Here, the relative pronoun *which* refers to *the government-run news agency*, which semantically is the source of the SIP *say*. However, the argument of *say* is *which*, and hence this will be the element selected as the source here. This case is parallel to example (26) above. Sentence (28) provides a second example of this.

(28) The move seemed aimed at heading off more trouble with Iran, **which**$_{c12}$ had condemned Iraq's invasion of Kuwait on Aug. 2 but also <u>criticized</u> the multinational force dispatched to Saudi Arabia.

• *The source is expressed by a phrase longer than one word-length.* If the source head is not marked in red, but another element of the source phrase is, select that element as the correct source.

(29) **They**$_{c1}$ don't want to play with us," one U.S. **crew**$_{c15}$ chief <u>said</u> .

9

The source of *said* in (29) is the whole phrase *one U.S. crew chief*, which has *chief* as its head. Since *chief* is its head, this is the element that should be marked in red (here, in bold face) as the source candidate of that SIP. But instead, the element that has been automatically selected as head is *crew*. We will consider this later element as the correct source of *said*.

II. The source is in fact a set of entities, expressed by means of a coordination structure. This is the case of the source of *objections* (figure 1, line #2), which actually includes *U.S.* as well as *British*, and the source of *filed* (figure 1, line #4), encompassing not only *U.S.* but also *British officials*.

When only one of the elements of the coordination is identified in red as candidate, we will select it as our choice (as in the examples above). On the other hand, if both are marked as candidates, we will select the first one.

III. The source does not correspond to any of the entities presented as candidates. That is, there is a segment in the sentence that expresses the SIP source, but neither its head nor any other of its constituents are marked in red.

Select the button labeled as OTHER; e.g., line #1 in figure 1, where the cognizer of *rejected* corresponds to *The World Court*.

IV. There is not an explicit segment referring to the SIP source. Select option NONE if no cognizer can be identified in the current sentence.

The following general guidelines can be of use for approaching each sentence:

Semantic criterion: First, be *semantically guided* in deciding what is the participant playing the role of SIP source.

Syntactic criterion: Once this participant has been identified, be *syntactically oriented*:

1. Select the source segment that is in a grammatical relation to the SIP (it is either one of its arguments or adjuncts). Personal and relative pronouns are therefore possible correct sources.

2. Identify the syntactic extent of the source (i.e., the syntactic phrase expressing it) and its head.

3. If the source is expressed by a phrase longer than one word, the head is ideally the segment that will be in red and needs to be selected. If however the element in red is not the head but another constituent in the phrase, select that constituent as the source.

4. If the source is expressed by a coordinated construction, apply the criterion in II above.

Option OTHER**:** If the source is not expressed by any of the element marked in red, select OTHER.

Option NONE**:** If the sentence has no explicit segment (word or phrase) expressing the source, select NONE.

3.3   Task 3: Annotating Factuality Assignments

The goal in this final task is selecting the factuality value that is assigned to each event by its relevant sources; in other words, deciding whether those sources evaluate the event as a fact that has happend, holds, or will happen for sure in the world; only as a possible fact; or as a counterfact. This task is fairly more complex than the previous two, so it is recommended that you review section 2 on *relevant notions* grounding the current research.

The annotation tool (figure 2) presents the sentences (third column) with the event to be assessed in blue. The relevant source is displayed in the fourth column, and the fifth column contains the factual values to select from.

Figure 2: Task 3 annotation screen

For some events, their factuality value is easily identifiable. Some others, however, pose a bit more of a problem since they may be embedded under other events or have several source chains that need to be evaluated against. The following sections provide 3 steps aimed at simplifying the annotation task for cases that are not clear. Step 1 and 2 are of a methodological nature and will help in complex cases. Step 3 is practical.

Throughout these annotation guidelines, the factuality value that a given source assigns to an event is formally expressed as: $f(event, source) = value$. For example, source author characterizing event e_2 as a fact in the world is presented as: $f(e_2, author)$=CT+.[4]

---

[4]The meaning of CT+ and other factual value abbreviations will be presented in section 3.3.3.

3.3.1   Step 1: What eventuality?

Goal: Identify the full event that needs to be assessed in terms of its factuality.

1. First, determine what is the proposition, clause, or phrase that fully expresses the event in question. For example, the complete units for the event expressions in bold face in (30) are given in (31). As you can see, some eventualities are included as part of other eventualities (e.g., e_2 is part of e_1).

   (30) On the other hand, it's **turning**$_{e_1}$ out to be another very **bad**$_{e_2}$ financial week for Asia.

   (31) **turning (e_1):** *It's **turning** out to be another very bad financial week for Asia.*
         bad (e_2):   *Another very **bad** financial week for Asia.*

2. Secondly, normalize the event expression. The normalized version will describe the event at focus in a neutral way; that is, without negative particles, markers of modality, etc. Such normalized version will be used in Step 3, for evaluating the factuality of the event.

   Negative polarity.   If the event is expressed in a context of negative polarity, transform it into its correspondent positive version:

   (32) Original sentence: *He had no **message** from Baghdad.*
         Normalized version: *He had a **message** from Baghdad.*

   In the case of embedded predicates, this normalizing step is applied one embedding level at a time. Consider the following sentence:

   (33) Al-Dosakee never **regretted**$_{e_1}$ [not **leaving**$_{e_2}$ Baghdad].

   For evaluating the factuality status of event e_1, the normalizing step corresponds to that in (34), whereas for evaluating event e_2, it corresponds to that in (35).

   (34) Original sentence: *Al-Dosakee never **regretted**$_{e_1}$ [not leaving Baghdad].*
         Normalized version of e_1: *Al-Dosakee **regretted**$_{e_1}$ [not leaving Baghdad].*

   (35) Original sentence: *Al-Dosakee never regretted [not **leaving**$_{e_2}$ Baghdad].*
         Normalized version of e_2: *Al-Dosakee **left**$_{e_2}$ Baghdad.*

   The normalization process aims at avoiding wrong factuality evaluations such as the following one:

   (36) Original: *Al-Dosakee never **regretted**$_{e_1}$ not leaving Baghdad.*
         Relevant sources: `author`.
         Factual assignment: f(e_1,`author`) = CT+

In the example above, the annotator decided that source `author` evaluates event e_1 as a fact (CT+). The wrong judgement comes from taking the predicate expression, together with the negative polarity marker, as referring to the event in question. According to the annotator, it is true (or a fact, CT+) that *Al-Dosakee never regretted not leaving Baghdad.* Nonetheless, the correct annotation is the the one in (37), in which the event of *Al-Dosakee regretting not leaving Baghdad* is assessed as a counterfact (CT−) according to source `author`.

(37) Original: *Bangin Al-Dosakee never* **regretted**$_{e_1}$ *not leaving Baghdad.*
Relevant sources: `author`.
Factual assignment: f(e_1,`author`) = CT−

Note that this normalizing step is applied regardless of the predicate type. In the previous example, for instance, there were two events marked with negative polarity, the first of which was factive. Both events were transformed into their correspondent positive versions.[5]

Modal markers. If the eventuality is qualified by a modal marker of any sort (auxiliaries, adverbials like *likely, it is probable that*, verbs like *seem, appear*, etc.), express the event in a neutral way by removing that marker. If the marker is an auxiliary verb, add the tense that best expresses the temporal reference in the original sentence.

(38) a. Original: *They now <u>can</u>* **begin**$_{e_1}$ *a more productive relationship.*
b. Normalized e_1: *They now <u>will</u>* **begin**$_{e_1}$ *a more productive relationship.*

(39) a. Original: *Before, the president <u>could</u>* **call**$_{e_2}$ *up to 200,000 reservists for up to 180 days without seeking congressional approval.*
b. Normalized e_2: *The president* **call<u>ed</u>**$_{e_2}$ *up to 200,000 reservists ...*

Events expressed by untensed clauses. Find the subject of the event and make it explicit. Then, add tense to the main predication, converting the clause into a full sentence.

(40) a. Original: **Facing**$_{e_1}$ *U.S. and Arab troops at the Saudi border, Iraq sought peace on another front today.*
b. Normalized: *Iraq* **faced**$_{e_1}$ *U.S. and Arab troops at the Saudi border.*

(41) a. Original: *The Sikh families received checks as* **compensation**$_{e_1}$ *for the riots.*
b. Normalized: *The Sikh families were* **compensated**$_{e_1}$ *for the riots.*

Events expressed by a noun phrase. Normalize the expression with the construction: *'NP is a fact'*, where NP stands for the NP headed by the event-denoting noun.

(42) a. Original: *With [new* **construction** *under way], ...*
b. Normalized: *[(The) new* **construction** *under way] is a fact.*

---

[5]In the case of factive predicates, it can be argued that transforming negated predicates into their positive counterparts causes loosing their presuppositional effect. As just shown, however, the normalized version is only used to identify the event being referred to, not to determine its factuality value.

Note that differences in tense (i.e., *'NP is a fact'*, *NP will be a fact'*) can lead to different factual values. Consider the following sentence:

(43) *There is no short-term hope for a diplomatic* **solution** *to the gulf crisis at least until economic sanctions force Saddam to withdraw his army.*

Event *solution* can be considered as a counterfact when evaluated in the present (*'The solution is a fact'*), but as a possibility when evaluated as an event in the coming future (*'The solution will be a fact.'*).

In the current research, evaluating the factual nature of nominals will remain restricted to a present time reference.

Events expressed by adjuncts to nouns, such as adjectival phrases (AP), prepositional phrases (PP), relative clauses, appositions, etc.

Normalize the expression with the construction: *'the N was/is/will be AC'*, where N corresponds to the noun being modified, and AC stands for the adjunct construction that denotes the event.

(44) a. Original: *With new construction [***under** *way], ...*
      b. Normalized: *(The) new construction is [***under** *way].*

Note that sometimes the category of the element marked as the event (from the TimeML annotation) does not correspond to the category of the whole event expression. This is the case of event e_2 in (30) above. Even though the marked element is an adjective, we consider the event is expressed as an NP, and hence we will normalize it as: *Another very bad financial week for Asia is a fact.*

Events expressed in conditional constructions (*if...else...*). Separate each clause of the construction as independent propositions. In the case of the antecedent, remove the conditional marker (*if, when*, etc.).

(45) a.   Original: *If the heavy outflows* **continue**$_{e_1}$, *fund managers will most probably* **face**$_{e_2}$ *increasing pressure to sell.*
      b.   Normalized e_1: *The heavy outflows will* **continue**$_{e_1}$.
          Normalized e_2: *Fund managers will most probably* **face**$_{e_2}$ *increasing pressure to sell.*

Events expressed in interrogative clauses. Convert the expression into its corresponding declarative form. If it has a *WH* particle, substitute it with a generic referring pronoun.

(46) a. Original: *The question is who will Cubans* **believe**?
      b. Normalized: *Cubans will* **believe** *somebody.*

(47) a. Original: *Is it* **low** *enough?*
      b. Normalized: *It is* **low** *enough.*

3.3.2 Step 2: According to what source?

Goal: Identify the sources that are assessing the factuality of the event at focus.

In other words, identify the relevant sources for that event. Relevant sources can consist of several sources in a nesting relation (refer to section 2.3), so we conceive them as *relevant source chains* regardless of whether they are constituted by only one source (e.g., `author`) or more (e.g., `she_author`). For event e_3 in the example below, normalized in (48b), there are three participants that have something to say about its factuality status: the author, Intel, and the customer who discovered the flaws:[6]

(48) a. Original: <u>Intel</u> said$_{e_1}$ *that last week a* <u>customer</u> *discovered$_{e_2}$ two* **flaws**$_{e_3}$ *in its 80486 microprocessor chip's floating-point unit.*

     b.   Normalized e_3: *It is a fact that there are two* **flaws**$_{e_3}$ *in its 80486 microprocessor chip's floating-point unit.*

The author is involved by default in the assessment of all events in a text. *Intel* and the *customer* are relevant here because they are the sources introduced by the SIPs *said* and *discover*, respectively, which are in an embedding relationship with e_3. Due to the nesting relation among these sources, the final relevant source chains for e_3 are: `author`, `intel_author`, `customer_author`, and `customer_intel_author`.

In the annotation tool, the relevant source chains for each event will already be given to you in the fourth column –hence, you don't need to understand how they are obtained. Furthermore, if they contain expressions that are explicit in the original sentence, you will see them marked in red in the text (in the third column).

The goal of this step is **understanding** the relation between the different sources in the chain (when it has more than one) and between these source and the event at focus. That is, understanding what it means for the factuality of event e to be assessed by the source chain $s_x\_s_y\_..._s_z$. For that purpose, it is helpful to make explicit the nesting relation between each source in a chain. For instance, we can recognize the assessment relations between e_3 in (48b) and each of its relevant source chains, by phrasing them as follows:

(49) Source chain: Assessment relation:

     `author`   *The* **author** *thinks/considers/says that e_3.*

     `intel_author`   *According to the* **author**, **Intel** *thinks/considers/says that e_3.*

     `customer_author`   *According to the* **author**, *the* **customer** *thinks/says that e_3.*

     `customer_intel_author`   *According to the* **author**, **Intel** *is of the view that the* **customer** *thinks/considers/says that e_3.*

The role of nesting sources. In a source chain, the *main source* is the first member of the chain (e.g., `customer` in the chain `customer_intel_author`), and the *nesting sources* are the remaining ones. Note that the factuality of the event is evaluated based on the main source. However, nesting sources are also important.

Take for example sentence (48) above. Source `customer` will assess the factuality of event e_3 differently depending on its nesting sources. Nested by `intel_author`, it will evaluate e_3 as

---

[6] From here onwards, the examples provided will present event expressions in bold face, and the strings denoting relevant sources for that event will be underlined.

certainly true (CT+), since it is the case that the customer takes e_3 to be a fact in the world, according to what Intel says, based on what is reported by the author:

(50) f(e_3, `customer_intel_author`) = CT+
 i.e., *According to the* **author***, Intel *says the* **customer** *considers that it is a fact that there are two* **flaws**$_{e_3}$.

But embedded only under `author`, the factual value is underspecified (Uu). This is because the discovery of the flaws is reported by Intel, and therefore the author is uncommitted to it. In other words, if asked whether the customer discovered two flaws, the author can reply: *I don't know, this is what Intel says.* Since the author is uncommitted to the discovery of two flaws, he is also uncommitted to whether the customer thinks/considers/says that there are two flaws in the mentioned chip's floating-point unit, hence:

(51) f(e_3, `customer_author`) = Uu
 i.e., *According to the* **author***, the* **customer** *considers that it is a fact that there are two* **flaws**$_{e_3}$

We will come back to this in section 3.4, on how to annotate SIP-embedded predicates.

Atypical sources. Certain types of sources require further consideration:

- **Generic sources:**

  Some source chains have GEN as one of their constituents (e.g., `GEN_author`). GEN refers to a non-explicit generic source, which can be rephrased as *everybody* or *somebody*, among similar expressions. In the example below, such source is implicitly introduced by the SIP *became clear*; i.e., it became clear to everybody.

  (52) a. Original: *At one point, when it* **became**$_{e_1}$ *clear controllers could not* **contact**$_{e_2}$ *the plane, someone* **said**$_{e_3}$ *a prayer.*
   b. Normalized e_2: *Controllers* **contacted**$_{e_2}$ *the plane.*
   c. Relevant sources: `author, GEN_author`

- **Dummy sources:**

  Some sentences are presented between quotation marks. If they are part of a longer quoted fragment, they will only have quotation marks at the beginning or ending:

  (53) a. "There are no unsettled **problems**$_{e_1}$ anymore.
   b. We have **solved**$_{e_2}$ them all."

  Since the current annotation is applied at the sentence level, there is no explicit source for fragments like these above. However, it is clear that they are reported by somebody who is not the author and that is most probably introduced somewhere else in the text. A dummy source (abbreviated as DUM) is introduced in these cases. Hence, the relevant source chains for events e_1 and e_2 above are: `author` and `DUM_author`.

  You will also find some sentences without quotation marks whose events may have dummy sources in their relevant chains. This is because the sentence is the continuation of a quotation opened at a previous sentence –and which will be closed at a posterior sentence.

16

- **Correferring sources:**

  Some sources point to the same participant. Sometimes, two of the sources in a chain correfer. For instance, one of the relevant source chain for event e_3 in (54) is `she_Hanna_author`.

  (54) a. Original: <u>*Hanna*</u> acknowledged$_{e_1}$ <u>*she*</u> told$_{e_2}$ *police interrogators that she* **prayed**$_{e_3}$ *for him to have a heart attack and die.*

      b. Normalized e_3: *She* **prayed**$_{e_3}$ *for him to have a heart attack and die.*

  Source chain `she_Hanna_author` presents the assessment of source *she* about e_3, according to what source *Hanna* acknowledged, based on what the `author` reported.

  Strictly speaking, however, *she* refers to the same person identified as *Hanna*. In the chain, this will be represented by the equality symbol (=); e.g., `she=Hanna_author`.

3.3.3 Step 3. What factuality value?

Goal: Determine the stance of the relevant source with regard to the factuality of the event. You can think it as the task of identifying what the source's answer would be if asked whether it is the case that such event took or will take place in the world. This is the step where the annotation for the task is finally carried out.

Factuality values. The set of the factuality values is presented in Table 1, where *Committed* and *Uncommitted Values* are evaluating the source attitude towards the factuality of the event, while the values grouped under *Other Values* are indications that will help me in the final design of the specification scheme.

Table 1: Factuality values

Value	Descriptor	Use
		Committed Values
CT+	Certainly positive	According to the source, it is **certainly** the case that X.
PR+:	Probably positive	According to the source, it is **probably** the case that X.
PS+	Possibly positive	According to the source, it is **possibly** the case that X.
CT-	Certainly negative	According to the source, it is **certainly not** the case that X.
PR-	Probably negative	According to the source it is **probably not** the case that X.
PS-	possibly negative	According to the source it is **possibly not** the case that X.
		(Partially) Uncommitted Values
CTu	Certainly positive or negative	The source knows whether it is the case that X or that not X.
PRu	Probably positive or negative	The source knows whether it is probably the case that X or that not X.
PSu	Possibly positive or negative	The source knows whether it is possibly the case that X or that not X.
Uu	Fully underspecified	The source does not know what is the factual status of the event, or does not commit to it.
		Other Values
Other	Other	Covering the following two situations - A different value is required here (e.g., U+, U-). - The annotator does not know what value to assign.
NA	Non-applicable	The factuality nature of the eventuality cannot be evaluated.

Committed and uncommitted values express two different but complementary types of information: **epistemic modality** and **polarity**. Epistemic modality refers to the degree of certainty

of a given source about whether an event is (or will be) a fact in the world. In the current work, it is systematized into the following categories, expressed as the initial part of the factuality value tag (e.g., <u>CT</u>+, <u>U</u>u).

CT: The source is certain; i.e., the source thinks the event took (or will take) place.

PR: The source thinks it is probable the event took (will take) place.

PS: The source thinks it is possible the event took (will take) place.

U: The source is uncertain (doesn't know), or uncommitted (doesn't have or doesn't express a view).

The second parameter, event polarity, expresses whether the event is presented as positive (happening) or negative (not happening). It occupies the second part of factuality value tags:

+: The event is seen as (certainly/probably/possibly) happening (factual).

−: The event is seen as (certainly/probably/possibly) NOT happening (counterfactual).

u: The polarity of the event is unknown or uncommitted.

Uncommitted values can sometimes be referred to as underspecified values as well because they are used when the source presents the event with some degree of underspecification: partial (CTu, PRu, PSu) or total (Uu). Partial underspecification describes the factuality of events like e_2 in the context below, evaluated according to source `john_author`.

(55) Original: *John knows$_{e_1}$ whether Mary* **came**$_{e_2}$.
    Normalized e_2:   *Mary* **came**$_{e_2}$.
    Factual assignment: f(e_2, `john_author`) = CTu

John is totally certain about the factual nature of e_2 (CT); it is not clear, however, what is the polarity he assigns to it: does he consider that Mary came (+), or that she didn't come (−)? Hence, the polarity remains underspecified (u).

Finally, the fully uncommitted (or underspecified) value (Uu) is used when:

- The source does not know the factual status of the event (56a),

- The source is not aware of the possibility of the event (56b), or

- The source does not overtly commit to it (56c).

The following examples illustrate each of these situations for e_2 when evaluated by source `john_author`:

(56)   a. *John does not know$_{e_1}$ whether Mary* **came**$_{e_2}$.

     b. *John does not know$_{e_1}$ that Mary* **came**$_{e_2}$.

     c. *John knows$_{e_0}$ that Paul said$_{e_1}$ that Mary* **came**$_{e_2}$.

Choosing the correct factual value. To select the factuality value of each event, we will use both its *normalized expression*, obtained from Step 1, and the *original sentence* where it appears. The normalized version gives a neutral definition of the event, whereas the original sentence provides the event as presented by the relevant source(s).

Given the event expressed in the normalized version, decide whether, in the original sentence, the source is characterizing it as: certainly happening (CT+), not happening (CT−), possibly/probably happening (PS/PR+), possibly/probably not happening (PS/PR−), or if, alternatively, the source presents it under some degree of underspecification –partial (CTu, PRu, PSu) or total (Uu).

Take for example event e_2 in (57b). We evaluate whether it is a fact according to source `author` and based on the information provided by the original sentence (57a):

(57) a. Original: *Women, children and invalids will be permitted$_{e_1}$ to **leave**$_{e_2}$ Iraq.*
   b. Normalized e_2: *Women, children and invalids will **leave**$_{e_2}$ Iraq.*

The process of assessing the factuality of an event can be guided by expressing the normalized event as a question according to the relevant source. For example:

(58) According to source `author`, will women, children and invalids **leave**$_{e_2}$ Iraq?

Assumptions on the evaluation context. The factuality value of events will be evaluated applying the two assumptions that follow:

I. **The naïve assumption:** As readers, we can assign different degrees of reliability to different sources. for example, experts are generally taken as highly trustworthy when informing about their field of expertise (59). On the other hand, politicians we dislike, or countries whose policies are questionable from our perspective, appear as less reliable (60).

   (59) a. Original: <u>*Experts*</u> *say$_{e_1}$ China is not **able**$_{e_2}$ to contain bird flu.*
      b. Normalized e_2: *China is not **able**$_{e_2}$ to contain bird flu.*

   (60) a. Original: <u>*China*</u> *says$_{e_1}$ it is **able**$_{e_2}$ to contain bird flu.*
      b. Normalized e_2: *It (China) is **able**$_{e_2}$ to contain bird flu.*

   We will however assume that all sources have the same degree of reliability. Hence, in spite of the difference of opinion about the same event e_2 in examples (59)-(60), source `author` will believe both sources equally.

II. **Locally-based knowledge assumption:**

   When evaluating the factuality of a given event, try to base your assessment uniquely on the knowledge available in the sentence expressing the event. Do not use either (i.) your personal knowledge about what happened in the world, or (ii.) other knowledge from sentences surrounding the one at focus.

   The following sentence provides a good example of an event that can be easily (but uncorrectly) evaluated using knowledge from the previous context.

(61) a.   Original: *Iraq said it* **invaded**$_{e_1}$ *Kuwait because of disputes over oil and money.*

     b.   Normalized e_2:   *Iraq* **invaded**$_{e_1}$ *Kuwait because of disputes over oil and money.*

It appears at the end of a document discussing the possible ways out of the crisis initiated by Iraq's invasion to Kuwait. Hence, it seems reasonable to analyze that source `author` commits to event e_1 (Iraq's invasion of Kuwait) as a fact in the world (CT+).

(62) Factual assignment: f(e_1, `author`) = CT+

This is however a judgement extrapolated from the previous text in the article, and not directly derived from the meaning of the sentence itself. We see that by placing, in the very same context, a sentence referring to an event not mentioned in the previous context (as e_2 below). In this case, source `author` can be clearly appreciated as uncommitted:

(63) a.   Original: *Iraq said it* **deserved**$_{e_2}$ *Kuwait because of historical rights.*

     b.   Normalized e_2:   *Iraq* **deserved**$_{e_1}$ *Kuwait because of historical rights.*

     c.   Factual assignment: f(e_1, `author`) = Uu

Discriminatory tests. What follows provides some guidance in distinguishing among the different values. It is mainly focussed on determining the *epistemic modality* value. *Polarity* can be added after the former is established.

1. **Underspecification (U) versus different degrees of certainty (CT/PR/PS):**

   The underspecified value (U) must be selected in the following cases:

   - **Uncommitted source.** Some events are presented by a given source as being witnessed, affirmed, denied, or thought to hold by somebody else. This *somebody else* can be fully committing to the event, but the former source is not.

     In (64), the source Sanders (i.e., `sanders_author`) commits to e_2, but `author` remains uncommitted since she is only presenting Sanders' opinion. Example (65) illustrates an equivalent case with e_2 in a context of belief.

     (64) a.   Original: *<u>Sanders</u> said*$_{e_1}$ *he'd* **double**$_{e_2}$ *his money.*

            b.   Normalized e_2:   *Sanders will* **double**$_{e_2}$ *his money.*

            c.   Relevant sources: `author, sanders_author`.

            d.   Factual assignments: f(e_2, `author`) = Uu

                                     f(e_2, `sanders_author`) = CT+

     (65) a.   Original: *Many <u>experts</u> thought*$_{e_1}$ *it would not be* **modified**$_{e_2}$ *soon.*

            b.   Normalized e_2:   *It will be* **modified**$_{e_2}$ *soon.*

            c.   Relevant sources: `author, experts_author`.

            d.   Factual assignments: f(e_2, `author`) = Uu

                                     f(e_2, `experts_author`) = CT−

- **Ignorant source.** The source does not know what the factual nature of the event is (66), or does not know about the event (67). Events falling in this classification are generally presented embedded under the predicate *know* (or similar ones; e.g., *discover, remember*) used in a context of negative polarity.

  In the examples below, the ignorant source is properly `john_author`. In (66), source `author` also assesses the factuality of e_2 as underspecified (Uu), but not beacuse he is ignorant about it, but because he is uncommitted (cf. *Uncommitted sources* above).

  (66) a. Original: *John does not know$_{e_1}$ whether Mary **came**$_{e_2}$.*
       b. Normalized e_2:   *Mary **came**$_{e_2}$.*
       c. Factual assignments: f(e_2, `author`) = Uu
                                  f(e_2, `john_author`) = Uu

  (67) a. Original: *John does not know$_{e_1}$ that Mary **came**$_{e_2}$.*
       b. Normalized e_2:   *Mary **came**$_{e_2}$.*
       c. Factual assignments: f(e_2, `author`) = CT+
                                  f(e_2, `john_author`) = Uu

- **Prospective event.** Prospective events are those expressed in a context of wish, promise, plan, decision, order, among many others. They are so called because, if they take place, they do so at a point in time after the event embedding them; namely, the event expressing the wish, promise, plan, etc.

  Prospective events are generally presented without any judgement regarding its degree of certainty. Contrast sentence (68) with sentence (69). In both of them there is a reference to the same normalized event e_3. Nonetheless, e_3 in (69) is explicitly qualified as a possible fact, whereas in (68) it is not.

  (68) Original: *Iraq has agreed$_{e_1}$ to allow$_{e_2}$ Soviets in Kuwait to **leave**$_{e_3}$.*
       Normalized e_3:   *Soviets in Kuwait will **leave**$_{e_3}$.*
       Relevant sources: `author`.
       Factual assignment: f(e_3,`author`) = Uu

  (69) Original: *Soviets in Kuwait will possibly **leave**$_{e_3}$.*
       Normalized e_3:   *Soviets in Kuwait will **leave**$_{e_3}$.*
       Relevant sources: `author`.
       Factual assignment: f(e_3,`author`) = PS+

To confirm the uncommitted nature of a given event, the following *copredication test* can be applied:

Check whether it is possible to copredicate it with both a context of positive certainty (CT+) and a context of negative certainty (CT−). Sentence (68) can be continued by either fragment in (70) (the first one presenting e_3 as certain, CT+, and the second, as false, CT−), whereas (69) does not.

  (70) a. ... *They will take the plane tomorrow early in the morning.* (CT+)
       b. ... *However, most of them decided to remain there.* (CT−)

2. **Absolute certainty (CT) versus degrees of uncertainty (PR, PS):**

Eventualities presented as certain (CT) cannot at the same time be assessed as *possible* (PS) in a context of *opposite polarity.* In the examples below, the symbol # is used to express that there is some sort of semantic anomaly.

(71) a. *Hotels are only* **thirty** (CT+) *percent full.*
      b. #*... but it is possible that they aren't* (PS−).

(72) a. *Nobody* **believes** (CT−) *this anymore.*
      b. #*... but it is possible that somebopdy does* (PS+).

On the other hand, eventualities that are characterized with some degree of uncertainty (PS or PR) allow it:

(73) a. $\underline{I}$ think *it's not going to* **change**$_{e_2}$ (PR−) *for a couple of years.*
      b. *... but it* could *happen otherwise.* (PS +)

(74) a. *He* probably **died** (PR+) *within weeks or months of his capture.*
      b. *...but it is also possible that the kidnappers kept him alive for a while.* (PS-)

In (73), the source I_author characterizes e_2 as PR− by presenting it under the scope of the predicate *think* used in 1st person. The fragment in (73b) can be added without creating any semantic anomaly. A similar situation is presented in (74): the adverb *probably* is characterizing the event as PR+, and the additional fragment presents the possibility of things being otherwise.

3. **Probable (PR) versus possible (PS):**

Distinguishing between the two shades of uncertainty is not always easy. The following hints can help in the task.

- **Presence of factualty markers.** Use the markers of modality that are present in the context, if any.

  POSSIBILITY (PS): *possibly, it's possible, maybe, perhaps; may, might, could.*
  PROBABILITY (PR): *probably, likely, it's probable, it's likely.*

- **Copredication with PR in contexts of opposite modality.** As seen, both degrees of uncertainty (PS and PR) accept copredication with PS in a context of opposite polarity (cf. the test to distinguish between CT and PR/PS above). However, only the lowest degree of uncertainty (PS) accepts copredication with PR in a context of opposite polarity.

  (75) a. $\underline{I}$ think *it's not going to* **change**$_{e_2}$ (PR−) *for a couple of years.*
        b. #*... but it* probably *will.* (PR+)

  (76) a. *It may not* **change**$_{e_2}$ (PS−) *for a couple of years.*
        b. *... but it most* probably *will.* (PR+)

- **Highest degree of uncertainty (PR).** See if the highest degree of uncertainty is possible.

  (a) Qualify the event with the modifiers *most probably* (PR) and *possibly* (PS), and choose the one that preserves the original meaning of the sentence.

  (b) If the event accepts to be qualified by at least one of the constructions below while preserving the meaning, select the highest uncertainty degree; that is, PR.

   – *not only possible but also probable*
   – *possibly, if not likely*
   – *possibly, and in fact likely*

4. **Other Values:**

   Other: Select the option *Other* if *a.* it seems that a different combination value should be assigned (e.g., U+ or U−), or *b.* it is not clear what the value should be.

   NA: Select *NA* if it seems that the event cannot be assessed in terms of factuality.

Discriminatory tests in a table. Table 2 summarizes the different copredication tests presented above in order to identify the degree of epistemic modality of a given event. Recall that copredication tests consist of testing whether the source in question could continue the sentence with another fragment where the event is placed in a context with modality or polarity values different from those in the original sentence.

Table 2: Tests for discriminating epistemic modality degrees

	$CT_=$	$CTop$	$PRop$	$PSop$
U	ok	ok	ok	ok
PS	ok	#	ok	ok
PR	ok	#	#	ok
CT	ok	#	#	#

In the table, the resulting epistemic modality values assigned to events are listed in the rows, while the tests are presented in the columns, abbreviated as $EM_{subindex}$. EM expresses the epistemic modality value of the context to be copredicated to the original sentence, whereas *subindex* indicates its polarity: = means context of the same polarity, and *op*, context of opposite polarity. For example, given an event e presented under a context of negative polarity in its original sentence, test PR_{op} requires creating a new fragment in which e is used in a context where the modality degree is *probable* and the polarity is positive: PR+.[7]

(77) Original: *I* think *it's not going to* **change**$_{e_2}$.   (PR−)
     Testing e_2 with PR_{op}:   #... but it probably will.   (PR+)

---

[7] Note that test CT= is non-discriminative. It is included because, combined with CP*op*, it allows identifying U values from the rest.

3.4 Annotated examples

This final section provides annotated examples of some very specific cases.

3.4.1 Events in future tense

Events in future tense will be evaluated in the same way as other tensed events. Hence, its normalized version will keep the tense as in the original. When assessing its factuality, the question to ask is whether the source commits to that event as a fact in the future.

(78) Original: A lawsuit in Germany will **seek**$_{e_1}$ a criminal prosecution of the
  outgoing Defense Secretary.
  Normalized e_1:   Same as the original.
  Relevant sources: `author`.
  Factual assignment: f(e_1,`author`) = CT+

Events embedded under certain SIPs (e.g., *predict, forecast*) can be evaluated in a similar way as events presented in future tense:

(79) Original: <u>Scientists</u> *predict*$_{e_1}$ that invisibility will **be**$_{e_2}$ possible for any object.
  Normalized e_2:   *Invisibility will* **be**$_{e_2}$ *possible for any object.*
  Relevant sources: `author, scientists_author`.
  Factual assignment: f(e_1,`author`) = Uu
  f(e_1,`scientists_author`) = CT+

The source being the agent of the prediction (*scientists*; that is: `scientists_author`) commits to the embedded event, but the source reporting the prediction by the scientists (i.e., `author`) remains uncommitted. Differences in the factual commitment of sources triggered by SIPs will be addressed in more detail in section 3.4.4.

3.4.2 Prospective events

The notion of *prospective event* was already introduced. They are events embedded under predicates belonging to any of the classes listed below, among others.

- Volition: *want, wish, expect.*

- Commissive: *promise, commit, propose.*

- Imperative: *order*

- Planning: *plan, decide.*

- ...

Due to the selection restrictions that these predicates impose on their embedded clauses, prospective events tend to be expressed by *to-* or *ing*-clauses.

The relevant source of a prospective event remains uncommitted as to whether it will happen or not (Uu). This is proved by the fact that the event accepts to be copredicated with contexts of both positive and negative absolute certainty (CT+ and CT−), as is the case with event e_1 in (82). The sentence can be continued with both fragments in (83).

(80) Original: *The UN ordered$_{e_1}$ Iran to* **halt**$_{e_2}$ *its nuclear activities.*
Normalized e_2:   *Iran* **halted**$_{e_2}$ *its nuclear activities.*
Relevant sources: `author`.
Factual assignment: f(e_3, `author`) = Uu

(81) a. *... Iran stopped its nuclear agenda one month later.* (CT+)

   b. *... but Iran disregarded the order.* (CT−)

A second example:

(82) a.   Original: *So for Sanders, it's time to* **express**$_{e_1}$ *his opinion.*
   b.   Normalized e_1:   *Sanders will express his opinion.*
   c.   Relevant chain: `author`
   d.   Factual assignments: f(e_1,`author`) = Uu

(83) a. *... although we all know that won't happen.* (CT−)

   b. *... which he will do in the next meeting.* (CT+)

3.4.3   Temporal clauses

Temporal clauses are introduced by adverbials like *when, as soon as , until*, etc. In a temporal clause, the tense of the main event has an effect on its factual nature.

Past tense: The event is presupposed as a fact in the world.

(84) Original: *As soon as he* **arrived**$_{e_1}$ *at the institution, he was placed under guard.*
Normalized e_1:   *He* **arrived**$_{e_1}$ *at the institution.*
Factual assignment: f(e_1, `author`) = CT+

Present tense: The event is situated at a future point in time. As illustrated in the following two examples, the factual degree assigned by the relevant sources may vary in each case. Apply your best judgement to decide.

(85) Original: *As soon as he* **arrives**$_{e_1}$ *at the institution, he will be placed under guard.*
Normalized e_1:   *He will* **arrive**$_{e_1}$ *at the institution.*
Factual assignment: f(e_1, `author`) = CT+

(86) Original: *He won't be under control until he is* **committed**$_{e_2}$ *to an institution.*
Normalized e_1:   *He will be* **committed**$_{e_2}$ *to an institution.*
Factual assignment: f(e_2, `author`) = PS+

3.4.4   SIP-embedded events

They are events under contexts of report, belief, knowledge, inference, witnessing, etc., created by SIPs like *tell, see, think*, or *know*. The complexity of the annotation task depends on the number of embeddings.

One level of embedding. There are two events to assess: the SIP and its embedded event. Consider:

(87) $\underline{AT\&T}$ **said**$_{e_1}$ *it would* **double**$_{e_2}$ *its assets.*

Event e_1, a SIP, is assessed as shown below. That is, its relevant source (`author`) considers that the saying event did certainly happen (CT+).

(88) Normalized e_1: $\underline{AT\&T}$ **said**$_{e_1}$ *it would double*$_{e_2}$ *its assets.*
     Relevant source: `author`.
     Factual assignments: f(e_1,`author`) = CT+

The second event, e_2, is embedded under e_1. As an embedded event, it will have more than one relevant source chain: the text author (`author`), but also *AT&T* according to the previous established relevant source (i.e., `AT&T_author`).

(89) Normalized e_2:   *It will* **double**$_{e_2}$ *its assets.*
     Relevant sources: `author, AT&T_author`.
     Factual assignments: f(e_2, `author`) = Uu
                               f(e_2, `AT&T_author`) = CT+

Event e_2 is directly reported by *AT&T* according to what the author says. Hence, source `AT&T_author` is committed towards that event as being a fact (CT+). This is however not the case of source `author`: if asked whether AT&T will double its assets, he can say *I don't know, I'm reporting only what AT&T says.*

We can apply the test for uncommitted factuality (copredication with contexts of positive and negative certainty, CT+ and CT−) by checking whether both fragments in (90) could be an adequate continuation of (87), according to source `author`. We see that it is the case.

(90) a. *... and so it did.*

     b. *... but it didn't happen.*

Other types of SIPs can result in different distributions of the factual values. This is the case of events embedded under the so-called factive predicates: predicates expressing knowledge (*know, discover, remember*) or some sort of psychological reaction (*regret, be glad that*):

(91) Original: $\underline{AT\&T}$ **knew**$_{e_1}$ *it would double*$_{e_2}$ *its assets.*

The assessment of the main predicate proceeds in the same way as example (88):

(92) Normalized e_1: $\underline{AT\&T}$ **knew**$_{e_1}$ *it would double*$_{e_2}$ *its assets.*
     Relevant source: `author`.
     Factual assignments: f(e_1,`author`) = CT+

Note however that, due to the different semantics of *know*, the factual status of its embedded predicate will also be assessed as a fact (CT+) by the embedding source (`author`):

(93) Normalized e_2:   *It will* **double**$_{e_2}$ *its assets.*
     Relevant sources: `author, AT&T_author`.
     Factual assignments: f(e_2, `author`) = CT+
                               f(e_2, `AT&T_author`) = CT+

Interaction with polarity and modality particles also plays a role:

(94) Original: $\underline{AT\&T}$ did not $\mathbf{know}_{e_1}$ it would $double_{e_2}$ its assets.

(95) Normalized e_1: $\underline{AT\&T}$ $\mathbf{knew}_{e_1}$ it would $double_{e_2}$ its assets.
Relevant source: `author`.
Factual assignments: $f(e_1, \texttt{author}) = CT-$

(96) Normalized e_2:   It will $\mathbf{double}_{e_2}$ its assets.
Relevant sources: `author, AT&T_author`.
Factual assignments: $f(e_2, \texttt{author}) = CT+$
                     $f(e_2, \texttt{AT\&T\_author}) = Uu$

Several levels of embedding.   The previous examples contained only one level of embedding, but due to the recursive property of language, more levels are also possible. Additional layers of embedding can introduce new sources, and that makes the analysis a bit more complex.

(97) The $\underline{newspaper}$ $\mathbf{reported}_{e_0}$ that $\underline{AT\&T}$ $\mathbf{said}_{e_1}$ it would $\mathbf{double}_{e_2}$ its assets.

The source `author` is committing to event e_0 as a fact:

(98) Normalized e_0:   The $\underline{newspaper}$ $\mathbf{reported}_{e_0}$ that $\underline{AT\&T}$ $said_{e_1}$ it would $double_{e_2}$
                     its assets.
Relevant chains: `author`.
Factual assignments: $f(e_0, \texttt{author}) = CT+$

On the other hand, `author` is uncommitted with regards to the factuality of e_1, since it has not been reported by him but by a different source, the $newspaper$ (i.e., `newspaper_author`). This is in fact the same situation illustrated in (89).

(99) Normalized e_1: $\underline{AT\&T}$ $\mathbf{said}_{e_1}$ it would $double_{e_2}$ its assets.
Relevant sources: `author, newspaper_author`.
Factual assignments: $f(e_1, \texttt{author}) = Uu$
                     $f(e_1, \texttt{newspaper\_author}) = CT+$

In this case, there are 4 relevant source chains for event e_2 (It will $\mathbf{double}_{e_2}$ its assets). They are: `author, newspaper_author, AT&T_author, AT&T_newspaper_author`.

(100) Normalized e_2:   It will $\mathbf{double}_{e_2}$ its assets.
Relevant sources: `author, newspaper_author, AT&T_author, AT&T_newspaper_author`.

What are the factual values that each of these source chains assign to event e_2? We start by those chains that have as its first member the participant most immediately connected to the factual assessment of the event –here, $AT\&T$. There are two of them: `AT&T_author` and `AT&T_newspaper_author`.

One possibility is to consider both chains as equivalent, and hence return the same factual value. Since it is clear from the sentence that $AT\&T$ commits to e_2 as a fact in the future, the factual values assigned by these two chains would be:

(101) Factual assignments: $f(e_2, \texttt{AT\&T\_author}) = CT+$
                     $f(e_2, \texttt{AT\&T\_newspaper\_author}) = CT+$

There is however a difference between `AT&T_author` and `AT&T_newspaper_author`. The first chain expresses the commitment of $AT\&T$ according to what the author knows or thinks. Put in other words, it expresses the factual value that the author considers $AT\&T$ assigns to e_2. On the other hand, `AT&T_newspaper_author` expresses the factual value that the author considers that *newspaper* sustains the $AT\&T$ assigns to e_2. Let's analyze each case in detail:

- `AT&T_author`: Expressing the factual value that the author believes $AT\&T$ assigns to e_2.

  The author cannot say what this value is, since he is uncommitted about whether $AT\&T$ said anything about doubling its assets (event e_1) –as seen in (99), it was the *newspaper* that commited to that as a fact. Hence, the factual value assigned by source chain `AT&T_author` to e_2 is underspecified: f(e_2,`AT&T_author`)=Uu.

- `AT&T_newspaper_author`: Expressing the factual value which the author thinks that *newspaper* considers the $AT\&T$ assigns to e_2.

  We already agreed that $AT\&T$ commits to e_2 as a fact in the future. This is according to what the *newspaper* reported, which at the same time, was reported by the author. Hence, f(e_2,`AT&T_newspaper_author`)=CT+.

Now it's time to see what factual values are assigned to e_2 by the two remaining relevant source chains:

- `newspaper_author`: Expressing the factual value that the author believes *newspaper* assigns to e_2.

  Event e_2 was reported by $AT\&T$. Whether it will be a fact or not is a piece of information that the *newspaper* may not have access to. All it knows (at least, according to the evidence we have from the current sentence), is what AT&T reported.

  Hence, this source remains uncommitted: f(e_2,`newspaper_author`)=Uu.

- `author`: Expressing the factual value that the author assigns to e_2.

  The author was already uncommitted regarding the factuality of e_1. He then remains also uncommitted regarding e_2: f(e_2,`author`)=Uu.

For reference, the final analysis is as follows:

(102) Normalized e_2:   *It will* **double**$_{e_2}$ *its assets.*
       Relevant sources: `author, newspaper_author, AT&T_author, AT&T_newspaper_author`.
       Factual assignments: f(e_2, `author`) = Uu
                            f(e_2, `newspaper_author`) = Uu
                            f(e_2, `AT&T_author`) = Uu
                            f(e_2, `AT&T_newspaper_author`) = CT+

As we saw earlier, the use of other predicates (e.g., the so-called factive predicates, such as *know* or *discover*) may result in different distributions of committed and uncommitted values:

(103) *The* <u>*newspaper*</u> **discovered**$_{e_0}$ *that* <u>*AT&T*</u> **said**$_{e_1}$ *it would* **double**$_{e_2}$ *its assets.*

(104) Normalized e_0:   *The <u>newspaper</u>* **discovered**$_{e_0}$ *that <u>AT&T</u> said*$_{e_1}$ *it would*
                           double$_{e_2}$ *its assets.*
   Relevant chains: `author`.
   Factual assignments: $f(e_0, \texttt{author}) = $ CT+

(105) Normalized e_1:   *<u>AT&T</u>* **said**$_{e_1}$ *it would double*$_{e_2}$ *its assets.*
   Relevant chains: `author, newspaper_author`.
   Factual assignments: $f(e_1, \texttt{author}) = $ CT+
                           $f(e_1, \texttt{newspaper_author}) = $ CT+

(106) Normalized e_2:   *It will* **double**$_{e_2}$ *its assets.*
   Relevant chains: `author, newspaper_author, AT&T_author, AT&T_newspaper_author`.
   Factual assignments: $f(e_2, \texttt{author}) = $ Uu
                           $f(e_2, \texttt{newspaper_author}) = $ Uu
                           $f(e_2, \texttt{AT\&T_author}) = $ CT+
                           $f(e_2, \texttt{AT\&T_newspaper_author}) = $ CT+

And similarly, the interaction with particles of modality and polarity also plays a role:

(107) *The <u>newspaper</u> did* NOT **discover**$_{e_0}$ *that <u>AT&T</u>* **said**$_{e_1}$ *it would* **double**$_{e_2}$ *its assets.*

(108) Normalized e_0:   *The <u>newspaper</u>* **discovered**$_{e_0}$ *that <u>AT&T</u> said*$_{e_1}$ *it would double*$_{e_2}$
                           its assets.
   Relevant chains: `author`.
   Factual assignments: $f(e_0, \texttt{author}) = $ CT−

(109) Normalized e_1:   *<u>AT&T</u>* **said**$_{e_1}$ *it would double*$_{e_2}$ *its assets.*
   Relevant chains: `author, newspaper_author`.
   Factual assignments: $f(e_1, \texttt{author}) = $ CT+
                           $f(e_1, \texttt{newspaper_author}) = $ Uu

(110) Normalized e_2:   *It will* **double**$_{e_2}$ *its assets.*
   Relevant chains: `author, newspaper_author, AT&T_author, AT&T_newspaper_author`.
   Factual assignments: $f(e_2, \texttt{author}) = $ Uu
                           $f(e_2, \texttt{newspaper_author}) = $ Uu
                           $f(e_2, \texttt{AT\&T_author}) = $ CT+
                           $f(e_2, \texttt{AT\&T_newspaper_author}) = $ Uu

Embedded contexts such as those above are a fundamental part of the current research. In real text, sentences may be more complex than the ones presented here. Please, take your time to annotate them well.

3.4.5   Modalized events

Events can be modalized by different types of markers: modal auxiliaries (*may, can,* etc.), modal adverbs (*possibly, certainly,*...), or predicates like *seem* or *appear.* Here we focus on some of the issues that each of these kinds of markers may trigger.

Modal auxiliaries: They can lead to ambiguity. Sometimes they depict the event as a fact in the world (111); in some others, as a possible fact (112); and in others they present it in a completely uncommitted way (113).

(111) Original: *As Dr. Snyderman manipulated the endoscope, the surgeon* <u>*could*</u>
                                   finally **see**$_{e_1}$ *the tumor.*
      Normalized e_1:   *The surgeon finally* **saw**$_{e_1}$ *the tumor.*
      Factual assignments: f(e_1,author) = CT+

(112) Original: *In the current situation, foreigners* <u>*could*</u> **become**$_{e_2}$ *hostages.*
      Normalized e_2:   *In the current situation, foreigners will* **become**$_{e_2}$ *hostages.*
      Factual assignments: f(e_1,author) = PS+

(113) Original: *In the past, the president* <u>*could*</u> **call**$_{e_3}$ *up to 200,000 reservists*
                                   without seeking congressional approval.
      Normalized e_3:   *In the past, the president* **called**$_{e_3}$ *up to 200,000 reservists with-*
                                   out seeking congressional approval.
      Factual assignments: f(e_1,author) = Uu

The tests in Table 2 confirm that event e_1 in (111) is seen as certainly a fact (CT+):

(114) a. **Test CT$_=$:**   (CT+)   *... And sure he saw it. It was huge.*
      b. **Test CT$_{op}$:**   (CT−)   #*... But he didn't see it.*
      c. **Test PR$_{op}$:**   (PR−)   #*... But probably he didn't see it.*
      d. **Test PS$_{op}$:**   (PS−)   #*... But possibly he didn't see it.*

    The distinction between the two remaining cases is more subtle, but again, the tests proposed in the previous section justify the different factuality values in each case. The test set in (115) validates PS+ as the factual value for e_2 in (112), and the test set (116), the value Uu for e_3 in (113).

(115) a. **Test CT$_=$:**   (CT+)   *... and that is what will happen.*
      b. **Test CT$_{op}$:**   (CT−)   #*... but that won't ever happen.*
      c. **Test PR$_{op}$:**   (PR−)   *... but it won't probably happen.*
      d. **Test PS$_{op}$:**   (PS−)   *... but it's also possible it won't happen.*

(116) a. **Test CT$_=$:**   (CT+)   *... And he used that right three times.*
      b. **Test CT$_{op}$:**   (CT−)   *... But he never did.*

Verbal predicates: Events can also be modalized by means of verbal predicates such as *seem* or *appear*. In this case, both the modal predicate and the modalized event will be assessed in terms of their factuality:

(117) Original: *Saddam* <u>**appeared**</u>$_{e_1}$ *to* **accept**$_{e_2}$ *a border demarcation treaty.*

(118) Normalized e_1:   *It* **appears** *that Saddam accepted*$_{e_2}$ *a border demarcation treaty.*
      Factual assignments: f(e_1,author) = CT+

(119) Normalized e_2:   *Saddam* **accepted**$_{e_2}$ *a border demarcation treaty.*
      Factual assignments: f(e_1,author) = PR+

Another frequent predicate modalizing the event is *expect* used in passive voice.

(120) Original: *It is* **expected**$_{e_1}$ *that the mounting will be* **completed**$_{e_2}$ *in April 1887.*

As with the previous cases, both the predicate *expected* and its embedded event will be evaluated in terms of its factual nature. The evaluation of *expected* in sentence (120) is as follows:

(121) Normalized e_1: *It is* **expected**$_{e_1}$ *that the mounting will be completed*$_{e_2}$ *in April 1887.*
Factual assignments: f(e_1,author) = CT+

Now consider the embedded event. Note that *expected* introduces and additional generic source, causing the embedded complement to have two relevant sources: author and GEN_author. We will assume that, in general, the factual value assigned by these sources is PR+ in both cases, although specific sentences may require different choices. As always, use your best judgement based on the tests provided in the previous section.

(122) Normalized e_2: *The mounting will be* **completed**$_{e_2}$ *in April 1887.*
Factual assignments: f(e_2,author) = PR+
f(e_2,GEN_author) = PR+

Adverbials: Typical adverbs of epistemic modality (e.g., *possibly, probably, certainly*) explicitly express the degree of certainty of the event, and hence they offer no problem. There are however other adverbs of evidential nature, such as *reportedly* and *apparently*, which also have an effect on the factuality of events.

The adverb *reportedly* has a purely evidential interpretation, equivalent to attributing a proposition to somebody else (e.g., *somebody said that...*). Because of that, the source of the sentence will always be evaluated as uncommitted:

(123) Original: *Apple Geniuses are* _reportedly_ **unbricking**$_{e_1}$ *iPhones.*
Normalized e_1: *Apple Geniuses are* **unbricking**$_{e_1}$ *iPhones.*
Factual assignment: f(e_1, author) = Uu

The adverb *apparently*, on the other hand, adds also a nuance of epistemic modality on top of its evidential interpretation. That is, in addition to expressing that the source of the information is somebody else (evidential), it also conveys some degree of certainty (*it is possible/probable...*) by the text source.

(124) Original: *Apple Geniuses are* _apparently_ **unbricking**$_{e_1}$ *iPhones.*
Normalized e_1: *Apple Geniuses are* **unbricking**$_{e_1}$ *iPhones.*
Factual assignment: f(e_1, author) = PR+

3.4.6 Negation

Negated events can be expressed in several ways. The following list reviews the most common means.

Negating the predicate expressing the event:

(125) Original: *She did_n't_* **follow**$_{e_1}$ *the rules.*
Normalized e_1: *She* **followed**$_{e_1}$ *the rules.*
Factual assignment: f(e_1, author) = CT−

Negating the subject:

(126) Original: *Nobody* **followed**$_{e_1}$ *the rules.*
  Normalized e_1: *Somebody* **followed**$_{e_1}$ *the rules.*
  Factual assignment: f(e_1, `author`) = CT−

Negating the object:

(127) Original: *She* **followed**$_{e_1}$ *no rules.*
  Normalized e_1: *She* **followed**$_{e_1}$ *the rules.*
  Factual assignment: f(e_1, `author`) = CT−

Embedded contexts provide additional strategies:

The negation is expressed as part of the lexical semantics of the embedding predicate:

(128) Original: *She failed to* **follow**$_{e_1}$ *the rules.*
  Normalized e_1: *She* **followed**$_{e_1}$ *the rules.*
  Factual assignment: f(e_1, `author`) = CT−

The embedding predicate is negated: This phenomenon applies to a restricted group of SIPs (*think, believe, expect*, etc.). A negation on the main predicate (the SIP) actually negates the event in the complement proposition.

(129) Original: *He does not* **think**$_{e_0}$ *[she* **followed**$_{e_1}$ *the rules].*

The annotation involves both the event expressed by the SIP (130) and the one in its complement (131).

(130) Original: *He does not* **think**$_{e_0}$ *[she followed*$_{e_1}$ *the rules].*
  Normalized e_0: *He* **thinks**$_{e_0}$ *she followed*$_{e_1}$ *the rules.*
  Factual assignment: f(e_0, `author`) = CT−

(131) Original: *He does not think*$_{e_0}$ *[she* **followed**$_{e_1}$ *the rules].*
  Normalized e_1: *She* **followed**$_{e_1}$ *the rules.*
  Factual assignment: f(e_1, `author`) = Uu
      f(e_1, `he_author`) = PR−

3.4.7 Hypothetical constructions

By hypothetical constructions I refer to constructions involving 2 events: a first one, that is presented as absolutely uncertain (Uu), and a second one, whose possibility of becoming a fact depends on the first one being a fact as well. The prototypical structures following this pattern are ***if... else...*** conditional constructions. In the example below, only the relevant events are marked:

(132) Original: *If the heavy outflows* **continue**$_{e_1}$*, fund managers will most probably* **face**$_{e_2}$ *increasing pressure to sell.*

(133) Normalized e_1: *The heavy outflows will* **continue**$_{e_1}$*.*
  Relevant chains: `author`
  Factual assignments: f(e_1, `author`) = Uu

Source `author`, the only relevant source here, is not asserting neither negating there will be a continuation of heavy outflows (e_1). Hence, f(e_1,`author`) = Uu. And since the factuality of e_1 is underspespecified, e_2 must be so as well. Note that this is the case even if that second clause has modality markers indicating a specific commitment of the source. Event e_2, for example, is qualified as *most probable*.

(134) Normalized e_2:   *Fund managers will most probably* **face**$_{e_2}$ *increasing pressure*
                                         to sell.
      Relevant chains: `author`
      Factual assignments: f(e_2,`author`) = Uu

Other hypothetical constructions are for example sentences with the structure: ***not... until/unless*** . In this case, the uncertain event is the one in the *until/unless* clause, and the dependent one, the one in the first clause. The same factuality analysis as for *if... else...* constructions holds:

(135) Original: *He wo<u>n't</u>* **do**$_{e_1}$ *it <u>until</u> Bloomberg* **is**$_{e_2}$ *within eight points or so in the polls.*

(136) Normalized e_1:   *He will* **do**$_{e_1}$ *it.*
      Relevant chains: `author`
      Factual assignments: f(e_1,`author`) = Uu

(137) Normalized e_2:   *Bloomberg* **is**$_{e_2}$ *within eight points or so in the polls.*
      Relevant chains: `author`
      Factual assignments: f(e_2,`author`) = Uu

It is important to notice, however, that past tense can lead to different factual judgments:

(138) Original: *He did<u>n't</u>* **do**$_{e_1}$ *it <u>until</u> Bloomberg* **was**$_{e_2}$ *within eight points in the polls.*

(139) Normalized e_1:   *He* **did**$_{e_1}$ *it.*
      Relevant chains: `author`
      Factual assignments: f(e_1,`author`) = CT+

(140) Normalized e_2:   *Bloomberg* **was**$_{e_2}$ *within eight points in the polls.*
      Relevant chains: `author`
      Factual assignments: f(e_2,`author`) = CT+

3.4.8   Interrogative constructions

Interrogative constructions include both interrogative sentences (141a) and indirect interrogative clauses (141b).

(141)   a. *Why did Mr. Phillips* **invent**$_{e_1}$ *a new type of head for screws?*

      b. *John knows*$_{e_0}$ *[why Mr. Phillips* **invented**$_{e_1}$ *a new type of head for screws].*

In some cases, the factuality of the event can be assessed as a fact or possibility:

(142) Original: *Why did Mr. Phillips* **invent**$_{e_1}$ *a new type of head for screws?*
      Normalized e_2:   *Mr. Phillips* **invented**$_{e_1}$ *a new type of head for screws.*
      Relevant chains: `author`
      Factual assignments: f(e_2,`author`) = CT+

(143) Original: *John knows$_{e_0}$ why Mr. Phillips* **invented**$_{e_1}$ *a new type of head for screws?*

Normalized e_1:   *Mr. Phillips* **invented**$_{e_1}$ *a new type of head for screws.*

Relevant chains: `author, john_author`

Factual assignments: $f(e_2,\texttt{author}) = \text{CT}+$

$f(e_2,\texttt{john\_author}) = \text{CT}+$

In some other cases, however, the factuality remainse underspecified:

(144) Original: *What did the president* **know**$_{e_1}$ *before declaring the war.*

Normalized e_2:   *The president* **knew**$_{e_1}$ *something before declaring the war.*

Relevant chains: `author`

Factual assignments: $f(e_2,\texttt{author}) = \text{Uu}$

(145) Original: *The country wonders$_{e_0}$ what the president* **knew**$_{e_1}$ *before declaring the war?*

Normalized e_1:   *The president* **knew**$_{e_1}$ *something before declaring the war.*

Relevant chains: `author, country_author`

Factual assignments: $f(e_2,\texttt{author}) = \text{Uu}$

$f(e_2,\texttt{country\_author}) = \text{Uu}$