DEFT Anomaly Belief-Sentiment Annotation Guidelines

Version 1.2

Linguistic Data Consortium

September 14, 2015

Table of Contents

1   Version History

This is version 1.2. All prior versions are working drafts that were unpublished.

2   Introduction

The goal of this annotation is to annotate belief and sentiment. In particular, we would like to know whether a particular source entity has a committed belief in a specific relation or event that he or she mentions. We would also like to know whether the source has positive or negative sentiment toward these same relations and events, as well as toward any entities he or she mentions. Similarly, if the speaker/writer is reporting belief or sentiment held by another person, we would like to know who this person is, as well as the type of belief or sentiment being held by that person.

3   Context

Context is very important. First, a given relation or event can be annotated in many different manners depending on how it is used:

(1)   a. *You were arrested?*
       b. *I know you were arrested.*
       c. *It's possible John was arrested.*
       d. *I hope John was arrested.*

As we will see, the three *arrest* events above will all be annotated differently. Furthermore, the larger context is crucial. The annotator must read the whole sentence in context and understand its meaning based on the context. The context includes textual context, but also conversational context if a dialog is being annotated. An example of the importance of context is the distinction between different uses of modal verbs (see Section 9.4).

4   Belief

For belief, we annotate whether the SW intends the hearer/reader ("HR") to interpret a relation or event as something about which the SW has a strongly held belief, as something the SW believes but not strongly, as someone else's belief, or as something towards which the SW has no belief (or possibly an entirely different cognitive attitude, such as desire or intention). If the SW mentions a relation or event about which someone else holds a belief, we also annotate this holder as well as their belief. The following list spells out some basic principles we will follow in annotating belief.

- The nature of the proposition containing the relation or event (whether it is an opinion, a statement about the SW's interior state, a claim about the world, a claim about a past event, or a claim about a future event) does not make a difference in the annotation. For the purpose of this task, any relations or events contained in *Paris is in France*, *John will travel to Paris tomorrow*, and *I see John* are all alike in that, by making these statements, the SW is directing the HR to assume the SW is committed to a complete belief in them.

- We do not annotate the truth of any statement made by the SW: real-world (encyclopedic) truth is not relevant. That is, it does not matter whether the SW is correct in a particular belief, only that the statements made indicate belief in a relation or event, whether or not the existence of that relation or event is objectively true in the real world. For example, *Paris is in France* and *Paris is in Russia* would both be considered committed beliefs of the SW toward the relations between *Paris* and *France* and *Paris* and *Russia*, respectively.

- Similarly, we do not attempt to distinguish whether or not the SW is telling the truth about his beliefs. If the SW indicates by his statements that he intends the HR to understand that he is committed to a belief in a particular relation or event, then it does not matter that the SW may in fact hold some other belief. See Section 7.1 for details.

- Sarcasm should also be disregarded for the purposes of this annotation task. If a SW says *No, Paris is in Russia*, the relation should be annotated as a committed belief, even if it seems likely that the SW is being sarcastic and in fact does not hold this belief. See Section 7.2 for details.

We have four categories of belief:

- **Committed belief (CB)**: the SW indicates in this utterance that he or she is certain the relation or event is true. For example, *I know Mark and Sandra have eloped* or *The sun will rise again*.

- **Non-committed belief (NCB)**: the SW identifies the relation or event as something which he or she considers possibly but not necessarily true; he or she does not have a strong belief in it. For example, *Mark and Sandra may have eloped* or *John may return tomorrow*.

- **Reported belief (ROB)**: the SW identifies the relation or event as something believed by someone else (be it committed or non-committed), but does not indicate his or her own degree of belief in the proposition. For example, *John said that Mark and Sandra have eloped*.

- **Not applicable (NA)**: the SW is not expressing any type of belief in the relation or event. This can be because the SW is expressing a different cognitive attitude, notably desire (*I want to leave* or *I wish Mark and Sandra would finally elope*) or intention (*I intend to leave*). However, it can also be that the SW expressly states that he or she has no belief in the relation or event: *I don't know whether Mark and Sandra eloped* makes clear that the SW thinks the *eloped* event is logically possible, but he or she does not know whether it is true.

In addition to the four categories of belief, we annotate the polarity of the belief:

- **Positive**: the SW indicates that the belief is about *p*.

- **Negative**: the SW indicates that the belief is about *not p*.

In the sentences *Paris is in France* and *Paris is not in France*, the SW has a committed belief toward to the relation between the entities *Paris* and *France*. The only difference in the beliefs is in their polarities, with the first being positive and the second being negative. It is important to

note that polarity is different from level of commitment to a belief. A doubt that an event happened (i.e. an NCB for belief with positive polarity) is different from a committed belief that the event did not happen (i.e. a CB with negative polarity). The specifics for annotation will be spelled out in Section 10.

5   Sentiment

For sentiment, we annotate whether the SW intends the HR to interpret their (or someone else's) attitude toward an entity, relation, or event as being positive or negative. As for belief, it does not matter if the SW is lying or being sarcastic. If there is linguistic evidence of a positive or negative sentiment toward an entity, relation, or event, we capture this by selecting one of the following three categories:

-   **Positive**: the SW expresses positive sentiment toward the entity, relation, or event. The evidence can be available in many different forms, such as a noun, modifier, or entire clause. For example, *I love John*, *The wonderful John stopped by*, *John is a life-saver*, and *I hope the judges choose John*, are all ways in which the SW can express positive sentiment toward the entity mention *John*.

-   **Negative**: the SW expresses negative sentiment toward the entity, relation, or event. Again, the evidence can be available in many different forms, such as a noun, modifier, or entire clause. For example, *I hate John*, *The annoying John just called*, *John is a disaster*, and *I hope the police arrest John*, are all ways in which the SW can express negative sentiment toward the entity mention *John*.

-   **None**: there is no linguistic evidence by the SW of either positive or negative sentiment toward the entity, relation, or event.

It is important to note that sentiment can vary between entities, relations, and events. The SW can express positive sentiment toward an entity, but negative sentiment toward an event in which the same entity is involved. For example, in the sentence *I'm disappointed that Mary cheated on the test*, the SW may have positive sentiment toward Mary, even though he or she is clearly expressing negative sentiment toward to *cheated* event. During annotation, we will be prompted to annotate entities separately from events and the entity mentions within those events, and so we must rely on context to determine which sentiment to attribute to each of these separately.

It is also important to not attribute your own sentiment, or the majority of people's sentiment, to an entity, relation, or event. Qualities or acts judged to be positive or negative by most people may be judged as the opposite by the SW. For example, if the SW states *Mary is a murderer*, you should use context to determine whether he or she views this quality positively or negatively, and not just assume a negative sentiment toward the entity *Mary* or toward any *murder* event in the surrounding context. In the event that the SW is happy that Mary is a murderer, we would want to capture that information.

6   Basic Concepts

As you have already seen, this document frequently uses the words *entity*, *relation*, and *event*. Although, for the purposes of this task, we will be given the entities, relations, and events in each document and do not have to locate them ourselves, it is useful to know what these terms mean in order to assign belief or sentiment to them.

6.1   Entities

An **entity** is a unique object or set of objects in the world – for instance, a specific person, place, or organization. A **mention** is a single occurrence of a name, nominal phrase, or pronominal phrase that refers to, or describes, a single entity. Multiple mentions of the same entity are said to be **coreferenced**.

Entities may be referenced in text at three different mention levels – a name, a common noun or noun phrase, or a pronoun. For example, the following are all mentions of the same entity occurring at different levels:

- Name Mention (NAM): Barack Obama
- Nominal Mention (NOM): the incumbent
- Pronoun Mentions (PRO): he, his

There are five entity types that we may come across:

- Person (PER) – Person entities are limited to humans. A PER entity may be a single person or a group.

- Organization (ORG) – Organization entities are corporations, agencies, and other groups of people defined by an established organizational structure. An ORG entity may be a single organization or a group.

- Geopolitical Entity (GPE) – GPE entities are composite entities, consisting of a physical location, a government, and a population. All three of these elements must be present for an entity to be tagged as a GPE. A GPE entity may be a single geopolitical entity or a group.

- Location (LOC) – Location entities are geographical entities such as geographical areas and landmasses, bodies of water, and geological formations as well as buildings and other permanent human-made structures. A LOC entity may be a single location or a group.

- Facility (FAC) – A facility is a functional, primarily man-made structure. Facilities are artifacts falling under the domains of architecture and civil engineering.

6.2   Relations

Relations are between ordered pairs of entities. The entities that participate in a relation are called **arguments**.

There are four **relation types** – Affiliation, Part-Whole, Physical, and Social – which comprise ten subtypes among them:

- Affiliation.Employment/Membership and Affiliation.Leadership

- Part-Whole.Subsidiary
- Physical.Located and Physical.Origin
- Social.Business, Social.Family, Social.Membership, Social.Unspecified, and Social.Role

For each relation, there is a trigger word, the two argument entities which are in relation, and the relation type and subtype.

6.3 Events

An **event** is something that happens at a particular place and time, involving specific participants. As with relations, the entities that participate in an event are called **arguments**, and are exhaustively tagged within the event mention scope. The date and place of the event's occurrence are also tagged arguments for events.

Events correspond to eight event types: Life, Movement, Business, Conflict, Contact, Personnel, Transaction, and Justice events. These eight types include a total of 33 subtypes among them:

- "Life" includes important events in a person's life: BE-BORN, MARRY, DIVORCE, INJURE, and DIE

- "Movement" presently includes only one subtype: TRANSPORT-PERSON

- "Business" includes important events in an organization's life-cycle: START, MERGE, DECLARE BANKRUPTCY, END

- "Conflict" includes: ATTACK and DEMONSTRATE

- "Contact" includes certain interpersonal communication events: MEET and COMMUNICATE

- "Personnel" includes events related to holding a job or other formal position: START, END, NOMINATE, ELECT

- "Transaction" includes two subtypes: TRANSFER-OWNERSHIP and TRANSFER-MONEY

- "Justice" includes important law-enforcement and judicial events and decisions: ARREST-JAIL, RELEASE-PAROLE, TRIAL-HEARING, CHARGE-INDICT, SUE, CONVICT, SENTENCE, FINE, EXECUTE, EXTRADITE, ACQUIT, APPEAL, and PARDON

For each event, we are given the event type and subtype, one or more of the arguments within the event mention scope, and a textual "trigger" – a string of text which specifically indicates the event. For the current annotation task, the event trigger is the unit that receives the belief and sentiment judgments.

7   Lying, Irony and Sarcasm

In this section, we discuss two communicative strategies that interfere with our basic annotation goal of annotating the private cognitive state of the SW. Lying and irony/sarcasm specifically introduce a deviation from the linguistic conventions normally used to signal a speaker's cognitive state. We are currently choosing to interpret the linguistic form literally, and thus ignore both lying and irony/sarcasm.

7.1 Lying

When the SW lies, he or she says an utterance conveying a proposition ("p"), believes that the proposition is not true ("not p"), and wants the HR to believe that the SW believes "p". We do not change our annotation if the SW appears to be lying: we take what the SW utters at face value. Real-world knowledge or even linguistic clues (whatever they may be) should not be used to attempt to detect lying. In this case, we actually do not attempt to detect the SW's true cognitive state with respect to belief, but rather his or her communicative intentions. Specifically, we annotate how the SW intends the HR to interpret a stated proposition with respect to the four categories of belief introduced in Section 2.

7.2 Irony and Sarcasm

When the SW uses irony/sarcasm, she says "p", believes "not p", but unlike the case of lying, she wants the HR to understand that she in fact believes "not p". Again, we do not change our annotation if the SW appears to be employing sarcasm: we again take the utterance at face value. In this case, we disregard the SW's communicative intentions and annotate how, ignoring irony/sarcasm, we assume the SW intends the HR to interpret a stated proposition with respect to the four categories of belief introduced in Section 2.

8 Basic Diagnostics for Belief

To judge a SW's belief in a relation or event, we can use the following diagnostics. To simplify the examples, we show only one relation or event for each sentence in order to make a point. It may be that the sentence contains multiple relations or events, but we can ignore that for now. We also do not show annotations for event arguments for simplicity. This type of annotation is showed in Section 10.

The following diagnostic tests make use of a linguistic concept called *entailment*. Proposition A is said to *entail* proposition B if, in order for A to be true, B must also be true. For example:

> A. *I took a plane to Turkey today.*
> B. *I traveled today.*
>
> A. *John F. Kennedy was assassinated.*
> B. *John F. Kennedy is dead.*

In each of the pairs of sentences above, if A is true, B must be true. If B turned out to be false, then logic dictates that A would also be false. In the examples that follow, the belief associated with the relation or event is shown after the word that triggers it, and both are in boldface.

- **Diagnostic 1: I am certain that…:** If the clause containing the relation or event can be embedded under *I am certain that…* (from the point of view of the SW or entity), then

that relation or event should be labeled **CB**. Another way of saying this is that the original sentence entails the newly formed sentence starting with *I am certain that…*. For example:

(2)   a. *John **traveled/CB** to Turkey.*
        b. *I am certain that John **traveled/CB** to Turkey.*

In (2) above, in some contexts, (a) *entails* (b). If the annotator determines that this is the case, the relation or event is labeled as **CB**.

Now consider a complex sentence:

(3)   a. *Mary knows that John **traveled/CB** to Turkey.*
        b. *I am certain that John **traveled/CB** to Turkey.*

We are here only interested in the *traveled* event, which should be CB on the part of the SW. (3b) is the test sentence created from the embedded clause of (3a). And in some contexts (a) entails (b), because the SW chose the verb *knows*: choosing the verb *knows* commits the SW to claiming belief over the thing that is known. The *traveled* trigger word should therefore be labeled CB. **Note that Mary also believes the *traveled* event to be true, but we do not need to annotate this unless the SW is only reporting Mary's belief without expressing a belief himself or herself.** To see this, consider this example:

(4)   a. *Mary thinks that John **traveled/ROB** to Turkey.*
        b. *I am certain that John **traveled/CB** to Turkey.*

Here, the choice of *thinks* is in opposition to the possible choice of *knows* – the SW is signaling that he or she is not telling us his or her own belief about the *traveled* event, only what Mary believes about it. In this case, we can only choose the **ROB** label.

Also note that this diagnostic sets a high bar, as does our definition of **CB**:

(5)   a. *I think that John **traveled/NCB** to Turkey.*
        b. *I am certain that John **traveled/CB** to Turkey.*

(5a) does not entail (5b), so *traveled* is not a CB.
Note that *believe* is a very complicated case. A devout Christian may say
(6a) and mean (6b):

(6)   a. *I believe that Jesus Christ is God's **son/CB**.*
        b. *I am certain that Jesus Christ is God's **son/CB**.*

However, when discussing distant relatives, (7a) does not seem to entail (7b); the choice of the verb *believe* suggests the SW is in fact not certain about the relation between Theo and Jesús.

(7)   a. *I believe that Jesús is Theo's **son/NCB**.*
        b. *I am certain that Jesús is Theo's **son/CB**.*

In spoken language, the two sentences may well have differences in intonation; in informal written language, there could be other signals such as changes in font or punctuation. We hope that this annotation will allow us to study these questions – the annotator should simply make judgments about what the SW or other entities believe; there is no need to record what linguistic elements signal the belief. (For sentiment, however, we do record this proof of positive or negative sentiment).

- **Diagnostic 2: I am not sure but think that…:** If the clause containing the relation or event can be embedded under *I am not sure but think that…* (from the point of view of the SW or entity), then that relation or event should be labeled **NCB**. For example:

(8) a. *Probably, John **traveled/NCB** to Turkey.*
      b. *I am not sure but think that John **traveled/NCB** to Turkey.*

In (8) above, in some contexts, (a) expresses the same as (b). If determined that this is the case in this context, label the *traveled* event as **NCB**. (See the Appendix for a list of phrases that may make a judgment NCB.)

- **Diagnostic 3: I have no opinion on this, but someone claims that…:** If the clause containing the relation or event can be embedded under *I have no opinion on this, but someone claims that…* (from the point of view of the SW or entity), then that relation or event should be labeled **ROB**. For example:

(9) a. *The AP reported that Iraq was **bombed/ROB**.*
      b. *I have no opinion on this, but someone (AP) claims that Iraq was **bombed/ROB**.*

In (9) above, in some contexts, (a) expresses the same as (b). If deter-mined that this is the case in this context, label the *bombed* event as **ROB**.

It is important to note that a relation or event cannot be labeled ROB if it is not a belief held by the reporter in the first place.

(10)   John said that Mary tried to contact/NA you.

In (10) above, the *contact* event is NA, because neither the SW nor John is expressing a belief about it. (See Section 9.6 for details on annotating matrix verbs of attempt.) John is only reporting that Mary tried. To test whether a relation or event is ROB or NA, think about how it would be annotated if the reporter were the SW. If John were the SW and said *Mary tried to contact you*, the *contact* event would be NA. Only beliefs held by the reporter (committed or non-committed) can be labeled ROB when reported.

- **Diagnostic 4: Matrix clause to adverb transformation**: If the matrix clause in which the relation or event is embedded has a first person subject, and if you can transform the matrix clause into an adverb which expresses some degree of certainty, then the relation or event should be labeled NCB or CB, depending on the degree of certainty expressed. For example:

(11) a. *I expect John to **travel/NCB** to Turkey soon.*
      b. *John will most probably **travel/NCB** to Turkey soon.*

In (11) above, in some contexts, (a) expresses the same as (b). If the annotator determines that this is the case in this context, label the *travel* event as **NCB**, because considering something to be probable but not certain indicates a non-committed stance.

(12) a. *I know that John will **travel/CB** to Turkey soon.*
  b. *John will most certainly **travel/CB** to Turkey soon.*

In (12) above, in some contexts, (a) expresses the same as (b), and clearly indicates the *travel* event is a **CB**.

(13) a. *I hope that John will **travel/NA** to Turkey soon.*
  b. *# John will probably/perhaps/maybe/certainly **travel/NA** to Turkey soon.*

In (13) above, the meaning of (a) cannot be conveyed by (b), so you cannot assume that the *travel* event is either **CB** or **NCB**. Instead, use other diagnostics; in this case, because the matrix verb is *hope* and its subject is the first person singular pronoun, indicating desire rather than belief, it should be marked **NA**.

9 A List of Constructions and Cases

9.1 Simplex and Matrix Clauses

A relation or event contained in an "unadorned" simplex (one-proposition) or matrix clause is usually CB on the part of the SW, even if (in the case of a matrix clause) the embedded clause expresses a different belief value. By "unadorned", we mean without adverbs, and not in a context which shifts the interpretation (e.g., a long passage of reported speech, a question). Examples of matrix clauses which are CB follow; the relation or event trigger in the matrix clause is underlined to distinguish it from other relations or events in the sentence.

(14) a. *Yeah, you **live/CB** in the district.*
  b. *I **send/CB** my son a check monthly and hope he **receives/NA** it.*
  c. *She **said/CB** she **heard/ROB** there's gonna be a big **party/NA**.*

In (14), the SW is certain about the *live*, *send*, and *said* events. The SW adds no additional information in these matrix clauses to lead us to believe otherwise.

9.2 Simple Questions

A relation or event in a question (*wh-* or *yes/no*) is always **NA**, since it is being queried, and therefore the SW clearly does not express any belief in its truth. For example:

(15) a. *Will John **travel/NA** to Turkey?*
  b. *Why would John **travel/NA** to Turkey?*

9.3 Imperatives

Some relations or events contained within imperatives are NA because they express a desire rather than a statement in which the SW believes:

(16) *Kill/NA that man!*

In (16) above, the SW is not expressing a belief in any actual *kill* event.

9.4   Modal Auxiliary Verbs

Modal auxiliary verbs (*may*, *might*, *can*, *could*, *should*, *would*, *need to*, *ought to*, *have to*) generally have two principal readings: epistemic or deontic. Epistemic modals express the SW's degree of certainty in a proposition, while deontic modals express obligation, permission, or ability. This section explains how to determine which reading is appropriate in a given context and how to annotate relations and events when these verbs are present.

- If used epistemically (as describing a non-certain belief), the modal auxiliary articulates the SW's cognitive attitude towards a single proposition about the relation or event. *Might* and *would* are always epistemic. *Might* is treated as described in this section, but for *would*, there are other considerations (see Section 9.12 on conditionals).

  A relation or event in a sentence containing an epistemic modal is most likely annotated NCB, since the modal generally indicates a lack of certainty.

- If used deontically (as describing or creating an obligation, permission, or ability), the modal is considered to introduce a complex proposition. In this case, there are two propositions: the proposition about obligation, permission, or ability, and the proposition that the obligation, permission or ability is about (i.e. the relation or event). The relation or event is annotated as NA, because the modal expresses the idea that it is possible, permitted, or required, but does not commit the speaker to a belief in its truth.

  (17) *John is so sick that he can't fly/NA to Turkey tomorrow.*

  One diagnostic for identifying a modal auxiliary as deontic is that it can be replaced by a construction using the verb *to be* + an adjective such as those shown in Table 1 below + the main verb of the proposition in the *to* infinitive form.

  (18) *John is so sick that he is not able to fly/NA to Turkey tomorrow.*

We note that, while it may be possible to relate these two meanings, for the sake of annotating belief, it is crucial to tease the meanings apart. The following list is not exhaustive, but is meant to show the type of reasoning that must be used to distinguish the epistemic and deontic meanings of modal auxiliaries.

- **Modal verbs of epistemic uncertainty**

  Most modals can be used to express non-committed belief of the complement clause.

Modal	Epistemic?	Deontic?	Deontic Adjective
may	yes	yes	allowed
might	yes	no	—
must	yes	yes	obliged
can	yes	yes	able
could	yes	yes	able
should	yes	yes	supposed

13

would	yes	no	—
need to	no	yes	required
ought to	yes	yes	—
have to	yes	yes	obliged

Table 1: Table of English modal auxiliaries with adjectival substitutions for deontic reading

(19) a. *John may be **traveling/NCB** to Turkey (but I am not certain).*
    b. *John could be **travelling/NCB** to Turkey (but I am not certain).*
    c. *John can't be **traveling/NCB** to Turkey (but I am not certain).*
    d. *John should have **arrived/NCB** by now (but I am not certain).*
    e. *John must have **arrived/NCB** by now (but I am not certain).*

In (19) above, the SW believes weakly in the *traveling* clause (or its negation, in the case of (19c)). There are meaning differences, notably that *can* and *could* relate more to deduction (possibility given other knowledge), while *may* more directly describes the SW's cognitive state. Note that *can* in general will be deontic rather than epistemic, while *can't* may be either deontic or epistemic.

As shown above, a diagnostic is to follow a construction with a modal auxiliary with *but I am not certain* – if the resulting sentence is plausible in context, the modal has an epistemic reading.

- **Modal verbs of obligation**

(20) a. *I need to **arrest/NA** you.*
    b. *I must **arrest/NA** you.*
    c. *I have to **arrest/NA** you.*

In (20) above, the SW surely believes that the *arrest* event must take place. However, the SW is not expressing a belief about the truth of the *arrest* event's existence, so it is labeled NA.

- **Modal verbs of permission or ability**

Permission is often expressed by *may*, as in the following example:

(21) *Students may **leave/NA** the compound at noon for up to two hours.*

Here, we interpret the sentence as describing a state of permission that affects students. However, what they may do is not asserted as having happened or not happened by the SW, and therefore the *leave* event is NA.

Ability is often expressed using *can*, as in the next example.

(22) *John can **jump/NA** 15 feet.*

The ability is asserted by the SW, but the SW is not asserting that the ability has actually been exercised (or will be).

Often, it is hard to distinguish ability from permission, but this is not necessary for belief annotation because both meanings are annotated the same way:

(23) *The prisoners can **leave/NA** now.*

Whether the prisoners are allowed to leave (because they have been re-leased) or able to leave (because they have found a way to escape), the SW has no belief about whether the *leave* event has or will happen.

- **Modal verbs of hope**

  There is another, less common use of *may* which expresses a wish or hope on the part of the SW.

  (24) May he **live/NA** long and **prosper/NA**!

  Because *may* is used to express a wish, the events *live* and *prosper* are NA.

Another test for determining whether a modal auxiliary is epistemic or deontic, is to add *I consider it possible that…* to the beginning of the sentence with and without the modal auxiliary. If the meaning is the same, it is an epistemic usage, otherwise it is a deontic usage. This can be combined with the other diagnostics discussed above to determine how to annotate a particular case. In (25) we apply three tests to determine how to annotate *John may arrive tomorrow*.

(25) a. *John may arrive tomorrow (but I am not certain this in fact will happen).*
      b. *# John is allowed to arrive tomorrow.*
      c. *I consider it possible that John may arrive tomorrow = I consider it possible that John will arrive tomorrow.*

(25a) seems natural, while (25b) does not (as indicated by the #). These results, combined with the fact that the two sentences in (25c) are equivalent in meaning, show that all the diagnostics point to this sentence being an epistemic usage. We therefore annotate it as follows:

(26) *John may **arrive/NCB** tomorrow.*

In (27) we perform the same tests on *Passengers may bring onboard at most one carry-on*.

(27) a. *# Passengers may bring onboard at most one carry-on (but I am not certain).*
      b. *Passengers are allowed to bring onboard at most one carry-on.*
      c. *I consider it possible that passengers may bring onboard at most one carry-on ≠ I consider it possible that passengers will bring onboard at most one carry-on.*

(27a) does not seem natural, but (27b) does; also, the two sentences in (27c) are not equivalent in meaning. Therefore, all the diagnostics point to this sentence being a deontic usage, and we annotate it as follows:

(28) *Passengers may **bring/NA** onboard at most one carry-on.*

9.5 Complex Questions

15

While the main proposition in a question (*wh-* or *yes/no*) is always NA (see Section 9.2), this does not mean that any embedded clauses are also automatically NA.

(29)   *Did you regret/NA giving/CB up your kingdom for a horse?*

In (29) above, the ***regret* event** is NA because it is being questioned (the SW is not expressing his or her own belief about it at all), while the *giving* event is CB, since the SW clearly believes that the HR did indeed give up his or her kingdom (otherwise, the SW would not be asking whether the HR regrets it). This issue is a case of presuppositions, which we discuss in further detail in Section 9.10.

Sometimes, questions are formed by adding what is called a tag question (e.g. *are you?*) to what would otherwise be a statement. The surrounding context of these types of questions should guide your annotation judgments.

(30)   a.   *John's **arriving/NCB** at 3, isn't he? I think he **said/NCB** that.*
       b.   *Sure his behavior is erratic, but you're not **seeing/CB** him **hurt/CB** anyone, are you?*

Although the examples in (30) are questions, none of the relations or events are labeled NA. This is because the SW does express a belief for each. In (a), the tag question only serves to weaken his or her belief in the *arriving* event, whereas in (b), the tag question does not affect his or her belief at all. The SW is certain that the man in question is not hurting anyone and that, in turn, the HR is not seeing anyone being hurt by him. While the form of the sentence is a question, it is a question to which the SW knows the answer.

9.6   Classes of Matrix Verbs

Here we discuss specific classes of matrix verbs.

- **Matrix verbs of desire, hope, and longing (DHL):**

  Relations or events that the SW wishes were true (past, present, or future) are marked as NA. Example: The complements of the verbs *hope* and *wish*. Note that negation does not change the rules for these verbs.

  (31)   a.   *I hope to **see/NA** you again soon.*
         b.   *I don't hope to ever **see/NA** you again.*

  In our example (31), the SW is clearly affirming his or her hope (or lack of hope), but the *see* event is NA regardless of whether the hoping is negated.

  It is important to note that a relation or event embedded under a matrix clause containing a DHL verb remains NA even when it is made clear that the desire has been fulfilled.

  (32)   *I'm glad we **met/CB** today, because I had hoped to **see/NA** you again soon.*

  In (32) above, the SW's wish has come true; however, the *see* event is not labeled CB, because, at the time of the wish, the SW was not expressing a belief about it.

- **Matrix verbs of expectation:**

If a relation or event is mentioned in the complement of a matrix verb that expresses an expectation, such as *expect* and *wait for*, it is annotated as NCB.

(33) *I expect Tom to **arrive/NCB** tomorrow.*

Because *expect* establishes a weak belief (as compared to the simple statement *Tom will arrive tomorrow*), we can see that *arrive* should be marked as NCB. **Expect should be marked CB, since the SW is clearly affirming that he holds some expectation.**

- **Matrix verbs of attempt:**

  The complements of verbs describing an attempt to carry out the action described in the embedded proposition are NA since the SW is not ex-pressing belief in success.

  (34) *you uh try to **lessen/NA** its impact*

9.7   Correlatives

These are constructions such as *the more we know, the better*. The SW has a committed belief about the correlation between *more* and *better*, but does not have a committed belief about *know*. Since there is no single word or contiguous string of words that express the correlation, we will annotate *better* as CB. *Know* will be annotated as NA:

(35) *The more we know/NA, the better/CB.*

9.8   Purpose Clauses

Verbs in purpose clauses are marked as NA. To determine if a clause is a purpose clause, replace *to* with *in order to*. If the utterance as a whole is still grammatical and does not change its meaning, the clause is a purpose clause.

(36) a. *I am **calling/CB** to **ask/NA** you for money.*
       b. *I am **calling/CB** in order to **ask/NA** you for money.*

In (36) above, *to ask you for money* is a purpose clause, as it could successfully be replaced by *in order to ask you for money*. Since this is the case, the *ask* event is labeled NA.

9.9   Relative Clauses

Relations and events in relative clauses are treated like others: the annotator makes a judgment about whether the SW believes in the existence of the relation or event mentioned in the relative clause.

(37) a. *Yesterday I **met/CB** Tom, who was at the **party/CB**.*
       b. *Tomorrow I may **meet/NCB** Tom, who will be at the **party/CB**.*

The relations and events in the main clause and the relative clause must be evaluated separately for the SW's belief. In (37a), the SW mentions a *met* event and a *party* relation, both of which she believes. In (37b), the SW is expressing a committed belief in the *party* relation

(i.e. she is certain that Tom will be there), but is not committed to the belief that she will meet him.

Restrictive relative clauses are used to identify referents. In written English, they usually do not have commas. Again, the annotator must make a judgment call about each relation and event separately.

(38) a. *The man who is **wearing/CB** a red shirt is my **brother/CB**.*
      b. *Any student who **fails/NA** the test will be **put/NA** on probation.*

In (38a), the SW is identifying a specific man and is making two statements about him: that he is wearing a red shirt, and that he is her brother. While the *wearing* **event** has the purpose to identify the correct referent, we can also suppose that the SW believes that the man in question is indeed wearing a red shirt. And clearly, the SW believes that the man is her brother. However, in (38b), there is no reason to assume that the SW believes that there actually are students who have failed or will fail the test – the sentence only expresses what happens should a student fail the test; it is quite possible that no student ever has and no student ever will fail. Therefore, the *put* event is also not expressing a belief, and both events are NA. The only CB being expressed in (38b) is that there is such a rule, but we do not annotate that currently. This is similar to the treatment of conditionals in Section 9.12, since (38b) could be reworded as *IF any student fails the test, he or she will be put on probation*.

Similarly, relations and events within relative clauses that modify general entity mentions, as opposed to specific ones, are to be treated as being hypothetical.

(39) a. *The government **helps/NA** anyone who **needs/NA** it.*
      b. *Friends who always **arrive/NA** late are annoying.*

Like the events in (38b), the events in (39a) are labeled NA. It may be the case that no one is in need of government help and that, in turn, the government helps no one. The quantifier *any* usually introduces such hypothetical scenarios. In (39b), the SW is not referring to a particular group of friends who always arrive late, and it may be the case that no friend always arrives late. The *arrive* event is thus labeled NA.

One final note about these types of clauses is that the belief associated with the relations and events can be changed entirely if commas are present.

(40) Friends, who always **arrive/CB** late, are annoying.

In (40), the presence of the commas demonstrates that the SW believes that all friends arrive late. The event within the relative clause is now CB.

9.10   A Note on Presuppositions in Complex Clauses

Presuppositions which are explicitly stated should be annotated as CB. The test for presupposition is that it is still asserted when the matrix clause of the sentence is negated. In the example below, the embedded proposition *eating junk food* is presupposed to be true no matter whether the matrix clause of the sentence is positive or negative:

(41) a. *You **stopped/CB eating/CB** junk food.*
      b. *You didn't **stop/CB eating/CB** junk food.*

9.11   If-Clauses: the Hypothetical & Conditional

Clauses with *if* are particularly complex; some express purely hypothetical conditions (in which case the result is necessarily hypothetical as well), as in the examples below.

(42) a. *If John **dies/NA**, I will **cry/NA**.*
      b. *If John **died/NA**, I would **cry/NA**.*

In (42) above, the SW has no belief in the *die* or *cry* events because they are purely hypothetical. So, both *die* and *cry* would be labeled NA.

In other cases, the condition (*if*-clause) has a known truth value (i.e., the speaker knows that it did or did not happen). This is typical of conditionals with the structure *if [noun] had [verbed]*, known as **counterfactual conditionals**, where the speaker is committed to the belief that the relation or event expressed in the *if*-clause did not happen.

(43) *If John had **died/CB**, I would have **cried/NA**.*

In (43) above, the SW knows that John did not die and now has a committed belief in the *died* event, which is labeled CB. The *cried* event, in contrast, should be labeled NA: in *if*-clauses, when the antecedent does not hold, we know nothing about the truth value of the result (the SW may have cried in any case).

(44) *I don't know if John **died/NA**, but if he were to **die/NA**, I would **cry/NA**.*

In (44) above, the SW has no belief in the *die* or *cry* events that follow *but* because he or she does not know if John did die. So, both events would be labeled NA. The first occurrence of *died* should also be labeled NA since the SW states he or she does not know if it is true or not.

Sometimes, the relationship of the propositions in conditional sentences can be more complex. Consider the following example.

(45) *If I doubted the **meeting/CB** were in Pittsburgh, I wouldn't have **flown/CB** to Pittsburgh.*

In (45), the SW has a committed belief in the *meeting* and *flown* events. He or she does not doubt there was a meeting in Pittsburgh because he or she flew to Pittsburgh (a case of negative polarity). The SW clearly believes that there is a meeting in Pittsburgh. Hence, the *meeting* event is CB. Also, the SW did in fact fly to Pittsburgh, so the *flown* event is also CB.

9.12   Temporal vs. Conditional Connectives

Clauses containing connectives, such as *once*, *unless*, *until*, and *when*, are also complex, since many of these connectives can be used either temporally or conditionally. If used temporally, to denote an event that occurred at a specific point in time, the proposition is usually CB.

(46) a. *I **got/CB** out of bed when I **heard/CB** my dog **bark/CB**.*
      b. *Once I **started/CB** **eating/CB** the chips, I couldn't **stop/NA**!*
      c. *I didn't **apologize/CB** until he **admitted/CB** he was wrong.*

In (46), most of the events are labeled CB, since it is clear that the SW is certain that the series of events occurred. However, many times the same connectives are used conditionally, sort of like *if*, and the events involved are labeled NA.

(47) a. *I **get/NA** out of bed when I **hear/NA** my dog **bark/NA**.*
    b. *Once I **start/NA eating/NA** chips, I can't **stop/NA**!*
    c. *I won't **apologize/NA** unless/until he **admits/NA** he was wrong.*

If you compare the examples in (47) to those in (46), you will notice that most of the events in (47) have not happened and may or may not happen in the future. In fact, (47a), for example, could be reworded as *IF I hear my dog bark, I get out of bed*. The events are therefore labeled NA. However, as with all judgments, these are based on context and are subject to change depending on the SW's choice of words. For example, in (48) below, the addition of the phrase *Every night* gives the conditional *when* of (47a) a more temporal reading.

(48) *Every night, I **get/CB** out of bed when I **hear/CB** my dog **bark/CB**.*

10   Annotating Belief

For each document, we begin by annotating any beliefs held towards the various relations and events mentioned throughout the document, as well as any arguments involved in the events. (We do not annotate arguments involved in the relations, because a belief about a relation is also a belief about the arguments involved.) In the tool, the Belief tab contains two subtabs: (1) Relations and (2) Events. It will be good to get in the habit of annotating all relations first, and then moving on to the Events tab and annotating all events.

10.1   Annotating Belief about Relations

In the Relations tab, there are three buttons that allow you to navigate through the relations contained within the document. For the current relation, a sentence representation of the relationship is displayed below these buttons, and the arguments involved are highlighted both here and in the document itself. Also underlined in the document are all of the entity mentions containing in the document.

Take, for example, the sentence in (49), which contains a relation triggered by the word *sister*.

(49) *He asked my sister to go to the party.*

A relation of this type (family) would be represented in the terms below:

(50) <u>my</u>   is related to   <u>my sister</u>

In (50), the relation shows the two arguments involved, which are both entity mentions. In this particular case, the entities involved are the entities co-refed with *my* (the SW) and *my sister* (the SW's sister). You'll notice that the entity mention closest to the relation head is the one that is displayed in this representation. For this task, you do not have to worry about selecting the nearest mention yourself, as these will already be provided. However, you should be aware that *my* is referring to the SW, whose name or username would also be underlined in the document.

In this example, you can also see why we do not annotate arguments of relations. Asserting that the SW has a committed belief in the "is related to" relation can only be done if the SW also is committed to the belief that *my* and *my sister* are the ones who are related.

Below the representation of the relation is the word that triggered the relation along with some information to help guide your judgments:

(51) *sister*
    - type: personalsocial
    - subtype: family

The exact wording of this information is not relevant for the present task; it is just meant to help you understand the basis of the relation. In (51), the information is fairly straightforward and may seem unnecessary. However, there will be several cases where clarification may be welcome. To the right of this trigger word is where we provide our three annotations:

- **Belief Type** – The options for this annotation correspond to the four categories described in Section 2. We mark whether the SW has a committed belief (CB) or non-committed belief (NCB) in the existence of the relation in question, or if the SW is reporting a belief (ROB) of another individual, or if the SW is asserting no belief at all (NA) about the relation.

- **Polarity** – If there is a belief (CB, NCB, or ROB), we mark whether it is about "p" or "not p" by selecting *positive* or *negative*, respectively. This information is gathered from context and is usually not contained in the trigger or arguments themselves.

- **Source** – If there is a belief, we mark the source of that belief. If we selected CB or NCB for Belief Type, the source will be the SW. If we selected ROB, the source will again be the SW, since he or she is the source of the report; however, selecting ROB will cause a new, identical row to appear. This is because we would like to know who exactly holds the belief and how strong the belief is. For the Source of this new line, we select the nearest mention of the entity that holds the belief that the SW is reporting. (See **Section 12** for more information on nesting.)

Below the trigger word section is a list of the arguments involved in the relation—though these arguments do not receive any annotations themselves—along with some information about them, as shown in the example below:

(52) *my*
    - role: per

    my sister
    - role: per

Again, (52) shows information that is very straightforward: both *my* and *my sister* are persons involved in the *sister* relation.

10.2   Annotating Belief about Events

The Events tab allows you to annotate any beliefs held about events mentioned throughout the document. This annotation is set up exactly like the annotation for relations, except that the arguments receive their own set of annotations.

In the sentence in (49) above (*He asked my sister to go to the party*), there is an *asked* event[1], which will be represented like:

(53) he   asked   my sister

The tool will then display the trigger word and accompanying information, next to which we provide our annotations:

(54) *asked*
        - type: contact
        - subtype: meet
        - realis: actual

The *realis* piece will come in handy when thinking about the level of belief to attribute to the event, though you should always verify this information beforehand[2]. Here, the *asked* event is something that actually happened. Since this information was gathered from the event's context, and the event's context was given by the SW, it is safe to assume that the SW is committed to the belief that the *asked* event did occur.

Information about the arguments is displayed below the trigger word:

(55) *he*
        - role: entity
        - realis: true

        my sister
        - role: entity
        - realis: true

This shows that both *he* and *my sister* are actual (as opposed to hypothetical) entities involved in the *asked* event. The SW makes it clear that both were involved, and so, for both, we annotate the belief type as **CB** and the polarity as **positive**.

Table 2 below shows the three events in (49) with their trigger words and arguments annotated for belief.

event1			event2		event3	
trigger	arg1	arg2	trigger	arg1	trigger	arg1

---

[1] In this particular sentence, there are actually three events in total, triggered by *asked*, *go*, and *party*. As we can see, more than just verbs trigger events.
[2] The information about relation and event trigger words and arguments that you see preloaded into the tool were extracted from another annotation task. They are meant to guide your annotations, but please note that, as with any annotation task, there may be some errors. When making a judgment about belief for this task, you should always verify that your judgments are appropriate given the context.

asked	he	my sister	go	my sister	party	my sister
CB	CB	CB	NA	NA	CB	NA

Table 2: Belief annotation for events in *He asked my sister to go to the party*.

The *go* event is labeled NA because, even though the SW is committed to the belief that his sister was asked to the party, he is not expressing a belief about whether or not she is going to the party. The *party* event is, however, labeled CB, because the SW makes it clear that he is certain the party exists. Since he is not expressing a belief in whether or not his sister will be at the party, the *my sister* argument is labeled NA.

As we can see, arguments must receive their own annotations for events, separate from the event trigger word. This is because, for example, someone can have a committed belief in the occurrence of some event, while being uncertain about who is involved in the event.

(56) *I think he asked my sister to go to the party.*

The sentence in (56) above could be annotated in several different ways depending on the context. For instance, the SW could be uncertain if the *asked* event happened, and so the trigger word and its arguments would be NCB:

(57) *I think he/NCB **asked/NCB** my sister/NCB to go to the party.*

Alternatively, the SW could be sure that *he* asked someone to go to the party but *thinks* that the person he asked is *my sister*. In this case, the event triggered by the word *asked* would be labeled CB, and the argument *my sister* would be labeled NCB:

(58) *I think he/CB **asked/CB** my sister/NCB to go to the party.*

It could also be the case that the SW is certain *he* asked *my sister* something, but is uncertain that the question involved going to the party. **This uncertainty would not be captured in this annotation task.**

Table 3 below summarizes the belief annotations over the sentence according to the various interpretations:

interpretation	event1			event2		event3	
	trigger	**arg1**	**arg2**	**trigger**	**arg1**	**trigger**	**arg1**
I'm not sure he *asked*.	asked	he	my sister	go	my sister	party	my sister
	NCB	NCB	NCB	NA	NA	CB	NA
I'm not sure *who* asked my sister.	asked	he	my sister	go	my sister	party	my sister
	CB	NCB	CB	NA	NA	CB	NA
I'm not sure *my sister* was the person he asked.	asked	he	my sister	go	my sister	party	my sister
	CB	CB	NCB	NA	NA	CB	NA

Table 3: Belief annotation for events in *I think he asked my sister to go to the party*.

11   Annotating Sentiment

After annotating for belief, we annotate any sentiment toward the various entities, relations, and events mentioned throughout the document. The relations and events in the Sentiment tab will be the same as those annotated in the Belief tab. We do not annotate sentiment toward arguments of relations or events. In the tool, the Sentiment tab contains three subtabs: (1) Entities, (2) Relations, and (3) Events. You should start by annotating all of the entities, and then moving on to the relations and then the events.

11.1   Annotating Sentiment toward Entities

The Entities tab contains three buttons that allow you to navigate through the entity mentions contained within the document. The word triggering the mention appears below these buttons and receives two annotations:

- **Polarity** – If there is a sentiment expressed to the particular entity mention, we mark whether it is *positive* or *negative*. If there is no sentiment, we mark *none*.

- **Source** – If there is sentiment, and we have marked either *positive* or *negative* under Polarity, then we mark the source of that belief.

Like belief, sentiment can be reported:

(59) *John said that Mary likes Tom.*

In (59), the entity mention *Tom* would receive positive sentiment, and the holder of this sentiment would be *Mary*. However, this is reported by *John*, and the entire sentence is being reported by the SW.

11.2   Annotating Sentiment toward Relations

The Relations tab within the Sentiment tab is set up much like the Relations tab within the Belief tab. The same sentence representation for each relation is present, but for each trigger word, there will be two annotations, as described in Section 11.1. As in the Belief tab, the arguments for relations do not receive their own annotations.

11.3   Annotating Sentiment toward Events

The Events tab within the Sentiment tab is set up much like the Events tab within the Belief tab. The same sentence representation for each event is present, but for each trigger word, there will be two annotations, as described in Section 11.1. Unlike for the Belief tab, the arguments for events here do not receive their own annotations.