

DEPARTMENT OF COMPUTER SCIENCE
University of Pittsburgh
Technical Report TR-02-101
Instructions for Annotating Opinions
in Newspaper Articles

by
Janyce Wiebe

Acknowledgement

This work was performed, in part, in support of the Northeast Regional Research Center (NRRC) which is sponsored by the Advanced Research and Development Activity (ARDA), a U.S. Government entity which sponsors and promotes research of import to the Intelligence Community which includes but is not limited to the CIA, DIA, NSA, NIMA, and NRO.

Contents

1	Introduction	3
2	Explicitly mentioned private states and speech events	4
2.1	Introduction	4
2.2	Nested sources of private states and speech events	4
2.3	The outside , inside , and on parts of sentences presenting private states and speech events	6
2.4	A private state or speech event should be annotated only if it is explicitly mentioned, not just inferred	13
2.5	Cases when a source (other than the writer) is not mentioned in the sentence.	14
2.6	The onlyfactive attribute: are opinions or other private states expressed? . .	16
2.7	Significant-and-particular private states and speech events	23
2.7.1	C1 - a significant portion of the contents of the private state or speech event must be given.	24
2.7.2	C2 - the private state or speech event must exist.	25
2.7.3	C3 - the noun phrase referring to the source (if there is one) must be referential	26
3	Expressive subjective elements	27
4	Type of attitude	30
5	Sentence-level classes	32
6	Strength	32
7	Certainty	32

1 Introduction

Picture an information analyst searching for opinions in the world press about a particular event. Our research goal is to help him or her find what they are looking for by automatically finding text segments expressing opinions, and organizing them in a useful way.

In order to develop a computer system to do this, we need people to annotate (mark up) texts with relevant properties, such as whether the language used is opinionated and whether someone expresses a negative attitude toward someone else.

Below are descriptions of the properties we want you to annotate. We will not give you formal criteria for identifying them. We don't know formal criteria for identifying them! We want you to use your human knowledge and intuition to identify the information. Our system will then look at your answers and try to figure out how it can make the same kinds of judgments itself.

This document presents the ideas behind the annotations.¹ A separate document will explain exactly what to annotate and how. An annotation system has been built using Sheffield University's Gate system (see <http://gate.ac.uk/> for information about Gate).

As of August 2002, the annotation system instructions and some sample annotations are available at
<http://www.cs.pitt.edu/mpqa/opinion-annotations/gate-instructions/>

When you annotate, please try to be as consistent as you can be. In addition, it is essential that you interpret sentences and words with respect to the context in which they appear. Don't take them out of context and think about what they *could* mean; judge them as they are being used in that particular sentence and document.

You will probably need to go through the instructions more than once.

As an overview, the annotations are centered around two main types of things: **(1) explicitly mentioned private states and speech events** (e.g., "John hates Bill" and "Mary said she would be home late"), and **(2) expressive subjective elements** (e.g., "to put it mildly" and "what an idiot"). When you annotate a text, you will identify text spans corresponding to these things, as well as other information, such as the source (e.g., whose opinion is being expressed?) and type (e.g., is the opinion a positive one?) Many of the same concepts are important for both (1) and (2). The instructions begin with explicitly mentioned private states (in section 2). By the time we get to expressive subjective elements (in section 3),

¹The conceptual framework used here is from (?; ?). It is built on work in linguistics and literary theory on *subjectivity*; that work directly addresses the linguistic expressions of attitudes. Please see (?; ?; ?; ?) for references. The framework of (?; ?) is an adaption of other work to the needs of NLP applications, by focusing on ambiguities that must be resolved to track point of view in language.

much of the conceptual framework needed has already been covered, so less space is required to explain them.

2 Explicitly mentioned private states and speech events

2.1 Introduction

We are particularly concerned in this project with texts that describe the *private states* and *speech events* of people.

A *private state* is an internal mental or emotional state; “private state” is a general term that covers opinions, beliefs, thoughts, feelings, emotions, goals, evaluations, and judgments. We can break a private state into components, viewing it as a state of an **experiencer** holding an **attitude**, optionally toward an **object**. For example, for the private state *John hates Mary*, the experiencer is *John* the attitude is *hate*, and the object is *Mary*.

A *speech event* is a speaking event, and includes writing as well as speaking. A speech event has a writer or speaker, obviously, and an object, namely, the thing that is written or said.

Sometimes, we will refer to the **object** of a private state or speech event as its **contents**. We mean the same thing by both.

For the sake of brevity, we will use one general term for **the experiencer of a private state, the writer of a speech event, or the speaker of a speech event**. **Any of these will be called a “source.”**

2.2 Nested sources of private states and speech events

Obviously, the writer of an article is a source: he wrote the sentences you are reading. But the writer may also write about other people’s private states and speech events, so we may have multiple sources in a sentence. For example, each of the following sentences has two sources: the writer (because he wrote the sentences), and Sue (because she is the source of a speech event in (1) and of private states in (2) and (3), namely thinking and being afraid).

(1) Sue said, “That’s a stupid idea.”

(2) Sue thinks that it’s a stupid idea.

(3) Sue is afraid to go outside.

Note, however, that we really don’t **know** what Sue says, thinks, or feels. All we know is what the writer tells us. Thus, (1), e.g., does not directly present Sue’s speech event but

rather Sue's speech event according to the writer. Thus, we have a natural nesting of sources in a sentence, and will view sentences as having levels. Sentences (1)-(3) have two levels each: a level for (the writer), and a level for (the writer, Sue).

There may be more than two levels, as in the following sentences:

(4) Sue said that Mary believes that Gore won the election.

Sources: (writer)(writer, Sue)(writer, Sue, Mary).

(5) Sue thinks that Bob said that the permits will no longer be accepted.

Sources: (writer)(writer, Sue)(writer, Sue, Bob).

(6) Sue is afraid that Daphne hates her.

Sources: (writer)(writer, Sue)(writer, Sue, Daphne).

How we might intuitively describe the levels of a sentence depends on the types of speech events and private states involved. As it happens, speech events are typically easier to describe than private states. Sentence (4), for example, is easy to describe, because it only involves speech events:

(writer) According to the writer, Sue said that Mary believes that Gore won the election.

(writer, Sue): According to the writer, according to Sue, Mary believes that Gore won the election.

(writer, Sue, Mary): According to the writer, according to Sue, according to Mary, Gore won the election.

Our intuitive description of sentence (6) is somewhat different, because it involves emotional private states (which are not well described with the phrase "according to"):

(writer): According to the writer, Sue is afraid that Daphne hates her.

(writer, Sue): According to the writer, there is a private state of Sue (that she is afraid that Daphne hates her).

(writer, Sue, Daphne): According to the writer, there is a private state of Sue toward a private state of Daphne (that she hates Sue).

Here are some real examples from newspaper articles:

(7) "Its aim of the 2001 report is to tarnish China's image and exert political pressure on the Chinese Government," human rights experts said at a seminar

held by the China Society for Study of Human Rights (CSSHR) on Friday.

Sources: (writer) (writer, human-rights-experts)(writer, human-rights-experts, report)

The entire sentence is attributed to the writer. The stuff in the quotes is attributed to (**writer, human-rights-experts**). The 3rd level, (**writer, human-rights-experts, report**), is introduced because a private state of the report is presented, namely that the report has the aim to tarnish China’s image and exert political pressure (according to the writer, according to the human rights experts).

(8) “It is heresy,” said Cao. “The ‘Shouters’ claim they are bigger than Jesus.”

Sources: (writer)(writer, Cao)(writer, Cao, Shouters)

Again, the entire sentence is attributed to the writer, and the stuff in quotes is attributed to (**writer, Cao**). At the (**writer, Cao, Shouters**) level, we have the Shouter’s claim that they are bigger than Jesus (according to the writer, according to Cao).

A note about **terminology**: (writer, Cao, Shouters) is the **nested source** of the claim speech event, and Shouters is its **immediate source**.

2.3 The outside, inside, and on parts of sentences presenting private states and speech events

Suppose a sentence presents a private state or speech event. The part of the sentence that presents the private state or speech event can be broken down into three pieces:

outside: The stuff outside the scope of the private-state or speech-event phrase (this includes, but is not necessarily limited to, the text span that designates the source).

on: The private-state or speech event phrase itself (for private states, this is the text span that designates the attitude).

inside: The stuff inside the scope of the private-state or speech-event phrase (this is the text span that designates the object).

Let’s consider an example sentence:

(15) John believes that Mary hates it that Tom believes that $2+2=5$

First consider **the writer's top-level speech event** (i.e., the writing of the sentence itself). The writer is implicitly the source, and the **outside** and **on** are implicit (i.e., something like, *The writer writes that* or *According to the writer*). The **inside** is the entire sentence.

- Now consider John's **believing** state:

The **source**: (writer, John)

The **text span** that presents it: "John believes that Mary hates it that Tom believes that $2+2=5$ ".

The **outside**: "John"

The **on**: "believes that"

The **inside**: "Mary hates it that Tom believes that $2+2=5$ "

- Next is Mary's **hatred** state:

The **source**: (writer, John, Mary)

The **text span**: "Mary hates it that Tom believes that $2+2=5$ "

The **outside**: "Mary"

The **on**: "hates it that"

The **inside**: "Tom believes that $2+2=5$ "

- Finally, we have Tom's **believing** state:

The **source**: (writer, John, Mary, Tom)

The **text span**: "Tom believes that $2+2=5$ "

The **outside**: "Tom"

The **on**: "believes that"

The **inside**: " $2+2=5$ "

Note that the **outside** may include more than a reference to the source. For example, consider the sentence:

“On Tuesday, John said that he was leaving, while hanging up the phone.”

The **outside** consists of “On Tuesday, John” and “while hanging up the phone.” The **on** is “said that” and the **inside** is “he was leaving”

There are different ways we could define the **outside** for a private state or speech event. We adopt the following **definition of an outside**: The **outside** for a particular private state or speech event P is simply the parent **inside** from the higher level minus P 's **on** and P 's **inside**. Because the **outside** is defined in terms of **ons** and **insides**, having a separate term for it is merely a convenience.

The remainder of this section presents a number of examples of **insides**, **outsides**, and **ons**. You may find that some of the **text spans** and **outsides** are non-intuitive. They are determined according to the definition of **outside** given just above. Don't worry too much about the **text spans** and **outsides**; the more important items are the **source**, the **ons**, and the **insides**.

(4) Sue said that Mary believes that Gore won the election.

• Sue's **speech** event:

The **source**: (writer, Sue)

The **text span**: “Sue said that Mary believes that Gore won the election”

The **outside**: “Sue”

The **on**: “said that”

The **inside**: “Mary believes that Gore won the election”

• Mary's **believing** state:

The **source**: (writer, Sue, Mary)

The **text span**: “Mary believes that Gore won the election”

The **outside**: “Mary”

The **on**: “believes that”

The **inside**: “Gore won the election”

(7) “Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government,” human rights experts said at a seminar held by the China Society for Study of Human Rights (CSSHR) on Friday.

- The human-rights-experts’ **speech** event:

The **source**: (writer, human-rights-experts)

The **text span**: “‘Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government,’ human rights experts said at a seminar held by the China Society for Study of Human Rights (CSSHR) on Friday.”

The **outside**: “human rights experts”, “at a seminar held by the China Society for Study of Human Rights (CSSHR) on Friday”

The **on**: “said”

The **inside**: “Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government”

- The report’s **aim**:

The **source**: (writer, human-rights-experts, report)

The **text span**: “Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government”

The **outside**: “Its”, “the 2001 report”

The **on**: “aim of”

The **inside**: “is to tarnish China’s image and exert political pressure on the Chinese Government”

(8) “It is heresy,” said Cao. “The ‘Shouters’ claim they are bigger than Jesus.”

- Cao’s **speech** event:

The **source**: (writer, Cao)

The **text span**: “‘It is heresy,’ said Cao. ‘The ‘Shouters’ claim they are bigger than Jesus.’”

The **outside**: “Cao”

The **on**: “said”

The **inside**: “It is heresy”; “The ‘Shouters’ claim they are bigger than Jesus.”

- The Shouters’ **claim**:

The **source**: (writer, Cao, Shouters)

The **text span**: “It is heresy”; “The ‘Shouters’ claim they are bigger than Jesus”

The **outside**: “It is heresy”; “The ‘Shouters’”

The **on**: “claim”

The **inside**: “they are bigger than Jesus”

(HR38.3) U.S. Secretary of State Colin Powell released the State Department’s annual report on human rights around the world yesterday, saying America’s concern for world freedom helps fight terrorism, instability, and conflict.

- Powell’s **speech** event:

The **source**: (writer, Powell)

The **text span**: “U.S. Secretary of State Colin Powell released the State Department’s annual report on human rights around the world yesterday, saying America’s concern for world freedom helps fight terrorism, instability, and conflict.”

The **outside**: “U.S. Secretary of State Colin Powell released the State Department’s annual report on human rights around the world yesterday”

The **on**: “saying”

The **inside**: “America’s concern for world freedom helps fight terrorism, instability, and conflict.”

- America’s **concern**:

The **source**: (writer, Powell, America)

The **text span**: “America’s concern for world freedom helps fight terrorism, instability, and conflict.”

The **outside**: “America’s”; “helps fight terrorism, instability, and conflict”

The **on**: “concern for”

The **inside**: “world freedom”

(**HR38.23**) Craner acknowledges that many people fear human rights issues will take a back seat to fighting terrorism in the Bush administration, but he says human rights are, in fact, a key part of that fight – a lesson, he suggested, that many in the White House had learned from bad experiences during the Cold War.

- Craner’s **acknowledgement**:

The **source**: (writer, Craner)

The **text span**: “Craner acknowledges that many people fear human rights issues will take a back seat to fighting terrorism in the Bush administration, but he says human rights are, in fact, a key part of that fight – a lesson, he suggested, that many in the White House had learned from bad experiences during the Cold War.”

The **outside**: “Craner”; “but he says human rights are, in fact, a key part of that fight – a lesson, he suggested, that many in the White House had learned from bad experiences during the Cold War.”

The **on**: “acknowledges that”

The **inside**: “many people fear human rights issues will take a back seat to fighting terrorism in the Bush administration”

- Many people’s **fear**:

The **source**: (writer, Craner, many people)

The **text span**: “many people fear human rights issues will take a back seat to fighting terrorism in the Bush administration”

The **outside**: “many people”

The **on**: “fear”

The **inside**: “human rights issues will take a back seat to fighting terrorism in the Bush administration”

- Craner’s **speech** event:

The **source**: (writer, Craner)

The **text span**: “Craner acknowledges that many people fear human rights issues will take a back seat to fighting terrorism in the Bush administration, but he says human rights are, in fact, a key part of that fight – a lesson, he suggested, that many in the White House had learned from bad experiences during the Cold War.”

The **outside**: “Craner acknowledges that many people fear human rights issues will take a back seat to fighting terrorism in the Bush administration, but he”; “a lesson, he suggested, that many in the White House had learned from bad experiences during the Cold War.”

The **on**: “says”

The **inside**: “human rights are, in fact, a key part of that fight”

- Craner’s **suggesting** event:

The **source**: (writer, Craner)

The **text span**: “Craner acknowledges that many people fear human rights issues will take a back seat to fighting terrorism in the Bush administration, but he says human rights are, in fact, a key part of that fight – a lesson, he suggested, that many in the White House had learned from bad experiences during the Cold War.”

The **outside**: “Craner acknowledges that many people fear human rights issues will take a back seat to fighting terrorism in the Bush administration, but he says human rights are, in fact, a key part of that fight”; “he”

The **on**: “suggested”

The **inside**: “a lesson”; “that many in the White House had learned from bad experiences during the Cold War”

- Many in the White House’s **learning** event:

The **source**: (writer, Craner, Many in the White House)

The **text span**: “a lesson”; “that many in the White House had learned from bad experiences during the Cold War”

The **outside**: “that many in the White House”; “from bad experiences during the Cold War”

The **on**: “had learned”

The **inside**: “a lesson”

- Many in the White House’s **bad experiences**:²

²**NOTE**: this private state does not pass criterion C1 (see section 2.7.1) for being a significant-and-particular private state (see section 2.7 below) – we do not see enough of the content of the experiences. To date, our annotators have been restricted to identifying only significant-and-particular private states, so they would **not** have annotated this private state.

The **source**: (writer, Craner, Many in the White House)

The **text span**: “a lesson”; “that many in the White House had learned from bad experiences during the Cold War”

The **outside**: “that many in the White House had learned from”; “during the Cold War”

The **on**: “bad experiences”

The **inside**: none

2.4 A private state or speech event should be annotated only if it is explicitly mentioned, not just inferred

You will be asked to identify explicitly mentioned private states and speech events, by identifying their **ons**, **sources**, and other attributes.

In this section, we want to highlight the word **explicitly**: the private state or speech event must be explicitly mentioned (or must be very intimately related to an action that is explicitly mentioned; see the next paragraph). It is not good enough that it may be inferred. For example, consider “John went to the store to buy some bread.” We can infer that he probably had the goal to buy bread and probably wanted bread, but please don’t annotate those kinds of inferred private states. Why not? Because almost all actions are intentional, so you would end up annotating almost all actions. We want to focus on private states and speech events in particular in this project.

What about the stuff in parentheses above: “(or must be very intimately related to an action that is explicitly mentioned)”? We call such actions **private-state actions**. Examples are: booing someone, shrugging ones shoulders, sighing heavily, shaking ones fist angrily, waving ones hand dismissively, smirking. **They are direct, physical expressions of private states; that is their reason for being.** Here is an example from a newspaper article:

(M1.19) As the long line of would-be voters marched in, snaking back and forth across the playing fields to save space, those near the front of the queue began to spontaneously applaud those who were far behind them.

- Those near the front of the queue’s **applauding** private-state action (the applause is a direct expression of a positive evaluation, and we do annotate it):

The **source**: (writer, those near the front of the queue)

The **text span**: “As the long line of would-be voters marched in, snaking back and forth across the playing fields to save space, those near the front of the queue began to spontaneously applaud those who were far behind them.”

The **outside**: “As the long line of would-be voters marched in, snaking back and forth across the playing fields to save space, those near the front of the queue”

The **on**: “began to spontaneously applaud”

The **inside**: “those who were far behind them”

In contrast, consider the following sentence from a newspaper article:

(9) After visiting the region several times, Italian senator Renzo Gubert, praised the Chinese Government’s efforts to protect the indigenous cultural and religious heritage while developing the local economy.

One might be tempted to annotate a private state for “the Chinese Government’s efforts to protect the indigenous cultural...” However, while one may infer a private state from “an effort to protect something,” we are being very stingy about annotating explicit private states and private-state actions. An **effort** is not literally a private state or speech event. Nor is it a private-state action, such as sighing, which is a direct physical expression of a private state. Thus, the only explicit private state or speech event annotated in (9) should be Gubert’s praise (whose **source** is (writer, Gubert) and whose **on** is “praised”).

2.5 Cases when a source (other than the writer) is not mentioned in the sentence.

This section describes a special situation that will arise occasionally when you annotate: a source (other than the writer) that should be annotated but is not mentioned in the sentence. There are two cases to consider.

The first case is when you need to fill in a previously-mentioned source in order to understand the sentence. Following are examples (the sentence(s) labeled **Context** are previous sentences included to provide context for the reader; the last sentence, e.g., **HR16.13** here, is the one we are discussing):

Context: The US anti-terrorist alliance has become the embodiment of two-sided justice.
(HR16.13): “If you’re not with us, you’re against us.”

- The US anti-terrorist alliance’s **speaking** event (the quoted speech is attributed by the writer to the US anti-terrorist alliance):

The **source**: (writer, US anti-terrorist alliance)

The **text span**: “If you’re not with us, you’re against us.”

The **outside**: none

The **on**: none

The **inside**: the same as the text span

Context: “We think this is an example of the United States using human rights as a pretext to interfere in other countries’ internal affairs,” Kong said. “We have repeatedly stressed that no double standard should be employed in the fight against terrorism,” he added. Beijing blames Uighur separatists for bombings and assassinations, and says they are linked directly to Osama bin Laden, the prime suspect in the September 11 attacks on New York and Washington.

(HR10.8) “By practicing double standard on human rights and minorities’ issues, the United States will definitely be despised and opposed by all righteous people in the world.”

- Kong’s **speaking** event (the implicit speaker is Kong):

The **source**: (writer, Kong)

The **text span**: “By practicing double standard on human rights and minorities’ issues, the United States will definitely be despised and opposed by all righteous people in the world.”

The **outside**: none

The **on**: none

The **inside**: the same as the text span

The second case is when there is a particular source, but it is never identified in the text. In this case, you should identify the **source** as **implicit**. Following is an example:

(HR37.6): U.S. officials have partially endorsed that view, pointing to rebel leaders such as Shamil Basayev and the Jordanian-born Omar Ibnal Khattab, who are believed to have

financial and other ties to Osama bin Laden’s al-Qaida terrorist network.

- The **believe** state (the source is implicit, i.e., the writer does not identify **who** believes the rebel leaders have financial ties to bin Laden’s terrorist network):

The **source**: (writer, implicit)

The **text span**: “U.S. officials have partially endorsed that view, pointing to rebel leaders such as Shamil Basayev and the Jordanian-born Omar Ibnal Khattab, who are believed to have financial and other ties to Osama bin Laden’s al-Qaida terrorist network.”

The **outside**: “U.S. officials have partially endorsed that view, pointing to”

The **on**: “who are believed to”

The **inside**: “rebel leaders such as Shamil Basayev and the Jordanian-born Omar Ibnal Khattab”; “have financial and other ties to Osama bin Laden’s al-Qaida terrorist network.”

2.6 The **onlyfactive** attribute: are opinions or other private states expressed?

Once you have identified the explicit private states and speech events in a sentence (and the associated information, such as their nested sources and their **ons**), the **onlyfactive** attribute may be added. You will classify each explicitly mentioned private state or speech event as either **onlyfactive=yes** or **onlyfactive=no**.

Suppose you have identified a private state or speech event P , and that you have decided its source is $(X1, X2, X3)$. The judgment we are making now is the following. According to $X1$, according to $X2$, is P a private state, or, in the case of a speech event, is P an expression of a private state? If P is a private state, then the tag is *onlyfactive=no*.³ If P is a speech event, then it could be either *onlyfactive=yes* or *onlyfactive=no*.

Sentences should always be judged in the context in which they appear.

³There is some question how to describe sentences such as “John knows that 2+2 is 4”. This is factive (according to the writer), but it also expresses something about the state of John’s knowledge (according to the writer). Note that this is a rare use of “know”. “Know” is more commonly used non-factively, as in “I just know that John will be late again!”. Please don’t worry about this case for now. Tag any cases you see as you think best, given the context in which the sentence appears. As we encounter instances of it, we’ll determine how to incorporate them into our overall conceptualization. One possibility is to define two subsets of private state, knowledge private states and the others, and say we are looking for the others. It is not clear to me, though, that we want to do this.

To decide if *onlyfactive* is yes or no, you will consider the **on** and **inside** for *P*. Sometimes, all you need to look at is the **on**: it may itself tell you that *P* is an opinion, emotion, claim, criticism, etc., making *onlyfactive* equal to no. Examples are “hates” and “criticizes”. If the **on** part is neutral, e.g., simply “said”, then you need to consider *P*’s **inside** and the context in which the sentence appears.

An easy example is, “John said, ‘Mary is an idiot’”, which is clearly an opinion at the level (writer, John). Thus, *onlyfactive* for John’s speaking event is no.

Note that, even if *onlyfactive=no*, *P* may involve something factual (according to X1, according to X2, according to X3). Consider the sentence “John criticized Mary for smoking.” The tag at the level (writer, John) is *onlyfactive=no*, because John’s negative evaluation of Mary is being expressed. However, this does **not** mean that John does not believe that Mary smokes.

We need to explicitly discuss an interesting case where the tag should be *onlyfactive=no*: facts and claims that are disputed in the context of the article. Consider “John said that smoking causes cancer.” In some articles, the tag at the level (*writer, John*) would be *onlyfactive=yes*. In other articles, it would be *onlyfactive=no*. You are not supposed to judge this with respect to your **own** beliefs, but instead consider the status of the proposition *smoking causes cancer* in the article. In a modern scientific article, it is likely to be treated as an undisputed fact. However, suppose it is an older article, in which the views of both scientists and tobacco executives are given about the relation of smoking to cancer. When the proposition is disputed, an *onlyfactive=no* tag should be used. Even if only the views of the scientists or only the views of the tobacco executives are explicitly given in the article, the *onlyfactive=no* tag might still be the right choice. **The tag *onlyfactive=no* would be the right choice if, for example, the scientists are arguing against the idea that smoking does not cause cancer. In this case, the scientists would be going beyond simply presenting something they believe is a fact, and arguing against an alternative view, and for the truthfulness of their own view.** We are very interested in these kinds of arguments in this research project, so we want you to identify them for us, tagging them *onlyfactive=no*.

Here is discussion of some of the example sentences we already talked about above. The same information (or a subset of it) is given as above, with the **onlyfactive** attribute and comments about it added:

(15) John believes that Mary hates it that Tom believes that $2+2=5$

- The writer’s **speaking** event:

The **source**: (writer)

The **text span**: the entire sentence

The **outside**: none

The **on**: none

The **inside**: the entire sentence

onlyfactive: yes.

Comments about onlyfactive: The **outside** and **on** parts are implicit, so we only have the **inside** part to consider. The **inside** part is that John believes something. That John believes that thing is presented as a fact at this level, so the tag is *onlyfactive=yes*.

• John's **believing** state:

The **source**: (writer, John)

The **text span** that presents it: “John believes that Mary hates it that Tom believes that $2+2=5$ ”.

The **outside**: “John”

The **on**: “believes that”

The **inside**: “Mary hates it that Tom believes that $2+2=5$ ”

onlyfactive: no

Comments about onlyfactive: The **on** phrase is *believes that*, which tells us that an opinion of John is being presented (according to the writer). You can see why we need to tease apart the various levels: *onlyfactive=no* does **not** mean that John does not believe the **inside** part; in fact, the sentence explicitly says he **does** believe it! The **onlyfactive=no** means that the writer is presenting John's opinion.

• Mary's **hatred** state:

The **source**: (writer, John, Mary)

The **text span**: “Mary hates it that Tom believes that $2+2=5$ ”

The **outside**: “Mary”

The **on**: “hates it that”

The **inside**: “Tom believes that $2+2=5$ ”

onlyfactive: no

Comments about onlyfactive: The **on** phrase refers to Mary's emotion/negative evaluation of something, so *onlyfactive=no*.

- Tom's **believing** state:

The **source**: (writer, John, Mary, Tom)

The **text span**: "Tom believes that $2+2=5$ "

The **outside**: "Tom"

The **on**: "believes that"

The **inside**: " $2+2=5$ "

onlyfactive: no

Comments about onlyfactive: This is a belief of Tom (according to the writer, ...), so the tag is *onlyfactive=no*.

(4) Sue said that Mary believes that Gore won the election.

- The writer's **speaking** event:

The **source**: (writer)

The **text span**: the entire sentence

The **outside**: none

The **on**: none

The **inside**: the entire sentence

onlyfactive: yes

- Sue's **speech** event:

The **source**: (writer, Sue)

The **text span**: "Sue said that Mary believes that Gore won the election"

The **outside**: "Sue"

The **on**: "said that"

The **inside**: "Mary believes that Gore won the election"

onlyfactive: yes

Comments about onlyfactive: The **on** part of the sentence is neutral: it is simply “said that”. There is nothing in this sentence that suggests that the **inside** part is Sue’s opinion. According to the writer, according to Sue, Mary does think that Gore won the election. Thus, the tag at this level is *onlyfactive=yes*.

- Mary’s **believing** state:

The **source:** (writer, Sue, Mary)

The **text span:** “Mary believes that Gore won the election”

The **outside:** “Mary”

The **on:** “believes that”

The **inside:** “Gore won the election”

onlyfactive: no

Comments about onlyfactive: The tag is *onlyfactive=no* because an opinion of Mary is presented (according to the writer, according to Sue).

(6) Sue is afraid that Daphne hates her.

- Writer’s **speaking** event:

The **source:** (writer)

onlyfactive: yes

- Sue’s **being afraid** state:

The **source:** (writer, Sue)

onlyfactive: no

- Daphne’s **hatred** state:

The **source:** (writer, Sue, Daphne)

onlyfactive: no

(7) “Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government,” human rights experts said at a seminar held by the China Society for Study of Human Rights (CSSHR) on Friday.

- Writer’s **speaking** event:

The **source**: (writer)

onlyfactive: yes

- The human-rights-experts' **speech** event:

The **source**: (writer, human-rights-experts)

The **text span**: “‘Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government,’ human rights experts said at a seminar held by the China Society for Study of Human Rights (CSSHR) on Friday.”

The **outside**: “human rights experts”, “at a seminar held by the China Society for Study of Human Rights (CSSHR) on Friday”

The **on**: “said”

The **inside**: “Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government”

onlyfactive: no

Comments about onlyfactive: The **on** for this saying event is “said”, which is neutral. Onlyfactive is no because, in the context of the article in which this sentence appears, it is argumentative to say that the aim of the report is to tarnish China’s image. This is an example of tagging a speech event **onlyfactive=no** because the claim is controversial or disputed in the discourse context of the article. In this article, the writer is arguing with what the report says, and is questioning its motives.

- The report’s **aim**:

The **source**: (writer, human-rights-experts, report)

The **text span**: “Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government”

The **outside**: “Its”, “the 2001 report”

The **on**: “aim of”

The **inside**: “is to tarnish China’s image and exert political pressure on the Chinese Government”

onlyfactive: no

Comments about onlyfactive: The **on** and **inside** segments are about a goal of the report. Goals are private states, so the tag is *onlyfactive=no*.

(8) “It is heresy,” said Cao. “The ‘Shouters’ claim they are bigger than Jesus.”

• Writer’s **speaking** event:

The **source**: (writer)

onlyfactive: yes

• Cao’s **speech** event:

The **source**: (writer, Cao)

The **text span**: “‘It is heresy,’ said Cao. ‘The ‘Shouters’ claim they are bigger than Jesus.’”

The **outside**: “Cao”

The **on**: “said”

The **inside**: “It is heresy”; “The ‘Shouters’ claim they are bigger than Jesus.”

onlyfactive: no

Comments on onlyfactive: The *onlyfactive* tag is no for two reasons. First, saying something is heresy is opinionated (unless in a technical religious-doctrine document). Second, opinionated language is used to characterize the “Shouters” (calling them “Shouters”, in quotes; “bigger than Jesus”). For a sentence about heresy to be judged *onlyfactive=yes*, the article would need to (1) simply assume the relevant religious tenets and laws, not also defend them against other possible views, and (2) the sentence corresponding to “The ‘Shouters’ claim they are bigger than Jesus” would need to be a non-sarcastic, technical presentation of the violation that constitutes the heresy.

• The Shouters’ **claim**:

The **source**: (writer, Cao, Shouters)

The **text span**: “It is heresy”; “The ‘Shouters’ claim they are bigger than Jesus”

The **outside**: “It is heresy”; “The ‘Shouters’”

The **on**: “claim”

The **inside**: “they are bigger than Jesus”

onlyfactive: no

Comments about onlyfactive:: The **on** is “claim”, which explicitly indicates that a claim of the ‘Shouters’ is presented (according to the writer, according to Cao), so the tag is *onlyfactive=no*.

(9) After visiting the region several times, Italian senator Renzo Gubert, praised the Chinese Government’s efforts to protect the indigenous cultural and religious heritage while developing the local economy.

- Writer’s **speech** event: **onlyfactive=yes**

- Gubert’s **praise** event:

The **source**: (writer, Gubert)

The **on**: “praise”

onlyfactive: no

Comments about onlyfactive:: “Praised” explicitly indicates that a positive evaluation was expressed (according to the writer), so the tag is *onlyfactive=no*.

You might ask yourself, **how strong does the private state have to be for onlyfactive to be no?** If you look at most sentences long enough, you may be able to perceive something opinionated, evaluative, etc. The **onlyfactive=no** tag should be used only for something significantly non-factive. Please use example annotated documents as a guide.

2.7 Significant-and-particular private states and speech events

Another important attribute of explicitly-mentioned private states is whether or not they are *significant-and-particular*. In fact, to date, the annotators have been asked to annotate only significant-and-particular ones. There are two reasons: (1) to make the annotation task more feasible, and (2) to identify clear, explicit expressions of private states and speech events rather than subtle minor ones, to support the NLP applications being developed.

Compatibility comment: As of August 2002, the accompanying instructions for performing the annotations do not refer to this attribute; they assume that the annotators are limiting themselves to significant-and-particular speech events and private states. This will be revised.

The expression of a private state or speech event P with source $(X1,X2,X3)$ is significant-and-particular only if the following three criteria are met:

- **C1** - a significant portion of the contents of the private state or speech event must be given
- **C2** - the private state or speech event must exist

- **C3** - the noun phrase referring to the source (if there is one) must be referential

Criteria **C1**, **C2**, and **C3** are described more fully in the following subsections.

Note that, unless it was explicitly indicated otherwise, all explicitly mentioned private states and speech events annotated so far in these instructions meet these criteria, and are significant-and-particular.

2.7.1 **C1 - a significant portion of the contents of the private state or speech event must be given.**

It is not sufficient that the private state or speech event be mentioned or described: we must be told something significant of the content of the speech event or private state in the sentence in which they are mentioned. We don't need to see the complete contents, but we do need some significant portion. Consider these examples from newspaper articles.

(10) As usual, the US State Department published its annual report on human rights practices in world countries last Monday.

Sources: (writer)

(11) And as usual, the portion about China contains little truth and many absurdities, exaggerations and fabrications.

Sources (writer)

In sentence (10), the report is referred to, and we know a topic it discusses, but that's not enough to satisfy criterion **C1**. We may be more tempted to say that the report's speech event in (11) meets criterion **C1**, because there are many descriptions of the contents of the report (little truth, many absurdities, exaggerations, and fabrications). However, we are not told **what** the report says, we are just told that the writer does not agree with it. So, criterion **C1** is not met in that case either.

Following are two examples of private states or speech events that do not meet criterion **C1** because they are merely referred to in the sentence. What makes these examples interesting is that the **ons** are noun phrases that refer back to a description of the private state or speech event in a previous sentence. However, to satisfy **C1**, we must be told something significant of the content of the private state or speech event within the current sentence.

Context: The US is planning to extend military aid to several countries in the alliance to enable them to root out terrorists without using US forces.

(**HR16.24**) Such wishful thinking risks making the US an accomplice in the destruction of human rights.

Comments: The private state whose **on** is "Such wishful thinking" does not meet criterion **C1**. This particular sentence does not tell us what the thinking is;

it just negatively evaluates it as “wishful”.

Context: It also says repression of the opposition, the media and civil society has worsened following Lukashenka’s re-election in September.

(HR38.15) President Islam Karimov of Uzbekistan takes similar criticism in the report, despite having become one of the key U.S. allies in the war on terrorism.

Comments: The **criticism** does not meet criterion **C1**. This particular sentence does not tell us what the criticisms are.

We end this subsection with an interesting example:

(12) The United States has violated the UN spirit of not interfering in other countries’ internal affairs,” said Jin Yongjian, a former deputy secretary-general of the United Nations.

The interesting private state is the one designated by “the UN spirit of not interfering in other countries’ internal affairs”. We decided that it **does** meet criterion **C1**, but we discussed the question at length when annotating this sentence. We decided this because **the spirit of not interfering in other countries’ internal affairs** is more like a private state than not: *spirit* in this sentence seems close to believing or valuing something. Plus, we see the contents of the private state: **not interfering in other countries’ internal affairs**. Some sentences, like this one, will require you to “roll the words around in your mouth a little” to decide which tag seems best.

2.7.2 C2 - the private state or speech event must exist.

The private state or speech event must exist (according to X1, according to X2). It must not be presented as merely hypothetical, for example.

Following are two examples with private states that do not meet criterion **C2**, namely Bush’s **belief** state in (HR16.22) and the governments’ **wanting** state in (HR38.28).

(**HR16.22**): If, however, US President George W. Bush still believes in the supremacy of human rights, he must not forget the breeding ground for terrorists that exists within the anti-terrorist alliance.

(**HR38.28**): “I’m also being pleasantly surprised by the number of governments – like I said, if they’re going to get their economies on a sound footing, and they

want a real relationship with us, they're going to have to start addressing these issues." (*said Craner* is implicit.)

2.7.3 C3 - the noun phrase referring to the source (if there is one) must be referential

Criterion **C3** is almost always satisfied; we include it just to rule out some weird cases. A private state or speech event P with source $(X1, X2, X3)$ meets criterion **C3** if the noun phrase referring to $X3$ is being used to refer to some particular person, group, or entity (according to $X1$, according to $X2$).

The following contain examples of private states that do not meet criterion **C3**:

(13) Any intelligent person would see that he is a jerk.

(HR16.16): No one who has ever studied realist political science will find this surprising.

Context: "We think this is an example of the United States using human rights as a pretext to interfere in other countries' internal affairs," Kong said. "We have repeatedly stressed that no double standard should be employed in the fight against terrorism," he added. Beijing blames Uighur separatists for bombings and assassinations, and says they are linked directly to Osama bin Laden, the prime suspect in the September 11 attacks on New York and Washington.

(HR10.8) "By practicing double standard on human rights and minorities' issues, the United States will definitely be despised and opposed by all righteous people in the world."

As used in these sentences, the phrases "Any intelligent person" in (13), "No one who has ever studied realist political science" in HR16.16, and "all righteous people in the world" in HR10.8 do not refer to specific groups or people. Thus, the corresponding private states do not meet criterion **C3**, namely, the **seeing** state in (13), the **find surprising** state in (HR16.16), and the **despising** and **opposing** states in (HR10.8).

As mentioned above, these cases are rare.

The source of a private state or speech event that *does* meet criterion **C3** may be quite vague, e.g., “critics say”, “westerners love their space”, “hunters and hikers support the bill”, etc. Here are sentences with examples:

(HR38.23) Craner acknowledges that many people fear human rights issues will take a back seat to fighting terrorism in the Bush administration, but he says human rights are, in fact, a key part of that fight – a lesson, he suggested, that many in the White House had learned from bad experiences during the Cold War.

Context: “I have been to Tibet many times. I have seen the truth there, which is very different from what some US politicians with ulterior motives have described,” said Xirao-Nima, who is a Tibetan.

(HR6.13) Some Westerners who have been there have also seen the ever-improving human rights in the Tibet Autonomous Region, he added.

(HR37.8) The Foreign Ministry said that the State Department pronouncement on Chechnya reflected the efforts of “certain circles” in the United States that object to the new U.S.-Russian partnership against international terrorism.

The phrases “many people” and “many in the White House” in (HR38.23), “Some Westerners” in (HR6.13), and “certain circles” in (HR37.8) are vague, but they still refer to groups of people that exist. Thus, they do not prevent their corresponding private states and speech events from meeting criterion **C3**.

3 Expressive subjective elements

Above, we focused on phrases that explicitly mention or describe private states and speech events, such as “Sue is afraid that ...”, and “Its aim of the 2001 report is to tarnish China’s image ...”

In this section, we consider words and phrases that **express** private states, but do **not** explicitly state or describe the private state. In other words, the **attitude** of the private state is implicit, and there is no explicit **on** for the private state. Such phrases, called **expressive subjective elements**, are often used by people to express their frustration, anger, wonder, negative evaluation, mirth, etc., without explicitly stating that they are frustrated, angry, etc. Sarcasm and irony often involve expressive subjective elements.

Examples of expressive subjective elements are: *what a, jerk*, and *!* in “what a jerk!”; *so-called* as in “so-called expert”; and *Far be it from me*, as in “Far be it from me to suggest

otherwise”. These express negative evaluation in these sentences. Other examples are *how* and *wonderful* in “How wonderful” and *absolutely* and *radiant* in “She was absolutely radiant.” These express positive evaluation in these sentences. As with explicitly mentioned private states and speech events, there are other types than positive and negative evaluation. An expressive subjective element might express surprise (e.g., *surprisingly*), it might express relief (e.g., *phew*), it might express obligation (e.g., *supposed to* as in “He was supposed to have finished the work hours ago”), etc.

Let me restate the specification of what you should look for, and look at some examples more closely:

- “What a jerk!”
The writer’s private state is expressed in this utterance, but is not explicitly mentioned. That is, there is no wording such as “I **judge** that person to be a jerk” (**judge** would be the relevant **on** phrase).
- “The so-called expert discussed the topic for 2 hours.”
The phrase “so-called” expresses the writer’s negative evaluation, suggesting that the person is not really an expert. Again, there is no wording such as “I do not think that he is an expert” (*to not think* would be the **on** phrase) or “In my opinion, that person is not an expert on that topic” (*opinion* would be the **on** phrase). Note that it is often difficult to come up with explicit paraphrases of the private states expressed by expressive subjective elements. That’s fine; you will not be asked to paraphrase them.
- “Far be it from me to suggest otherwise, but ...”
Again, there is no explicit word or phrase referring to the huffiness or indignation expressed by the phrase “far be it from me.”
- “John blathered on and on about his success on the racetrack.”
The phrase *blathered on and on* expresses the writer’s negative evaluation of John’s speaking events. Thus, it is an expressive subjective element with (writer) as the source.

Often, as in the above examples, the source of the private state is the writer. However, there may also be expressive subjective elements attributed to a nested source. An example is “heresy” in (8):

(8) “It is heresy,” said Cao. “The ‘Shouters’ claim they are bigger than Jesus.”

The **source** of the negative evaluation expressed by “heresy” in (8) is (**writer, Cao**).

Interestingly, an expressive subjective element may “skip levels”, as follows: let P be a private state or speech event with source $(X1, X2, X3)$. It is possible that, even if an expressive subjective element SE is in the **inside** for P , the source of SE could be $(X1, X2)$ or even

(X1). Consider the sentence:

(14) John thinks that that moron is brilliant.

The only way to make sense of (14), at least without further context, is that “that moron” is the writer’s description, not John’s. To match the above description: *P* is John’s **thinking** private state, whose **source** is (**writer, John**) and whose **inside** is “that moron is brilliant”. However, even though the expressive subjective element “that moron” is part of the **inside** of *P*, the source of “that moron” is (**writer**) (not (**writer, John**)).⁴

Sentence (8) contains an example:

(8) “It is heresy,” said Cao. “The ‘Shouters’ claim they are bigger than Jesus.”

The phrase “bigger than Jesus” is somewhat sarcastic (consider “bigger than Madonna” or “bigger than Elvis”). It is in the **inside** for (**writer, Cao, Shouters**)’s **claim**, but the source of the sarcasm expressed by “bigger than Jesus” is (**writer, Cao**).

Please make judgments concerning expressive subjective elements independently from your **onlyfactive** judgments.

There is one class of words and phrases that we will sometimes treat as expressive subjective elements, and sometimes as private-state terms, namely words and phrases such as *seem*, *appear*, and *give the impression that*. Consider how these terms may be used:

- With definite sources (treat as explicitly mentioned private states, not as expressive subjective elements):

It appeared to John that Mary was sick (“appeared” is treated as an **on**)

It seemed to John that Mary was sick (“seemed” is treated as an **on**)

John had the impression that mary was sick (“had the impression” is treated as an **on**)

- Without an explicit source (treat as expressive subjective elements):

⁴To use Banfield’s (?) terminology, subjective elements such as “that moron” are “embeddable”, in contrast to those such as “!” which are not.

It appears that Mary was sick (“appears” is treated as an expressive subjective element rather than an **on**)

(HR40.??) These rules seem to legitimize violence and terrorism one way or the other, provided that they target unarmed Palestinians or non-Israelis (“seem” is treated as an expressive subjective element rather than an **on**)

- Explicit “one” sources (treated as expressive subjective elements):

(HR37.2) “One gets the impression that its writers simply used old drafts, as if nothing had happened in either Russia or the United States in recent years, as if the events of Sept. 11, 2001 had not occurred and the international community had not closed ranks in the battle against terrorism,” the ministry said in a statement (“the impression” is treated as an **expressive subjective element** rather than as an **on**)

We can summarize the treatment of these terms as follows. They are treated as private-state terms only when there is an explicit, definite source. Otherwise, they are treated as expressive subjective elements.

4 Type of attitude

An important property of private states is the type of attitude they involve.

Recall that the annotations involve private states in three situations:

1. explicitly mentioned private states
2. explicitly mentioned speech events judged to be **onlyfactive=no**
3. expressive subjective elements

You may be asked to identify the type of the private state (or, more specifically, the type of the attitude) in all three cases.

For now, we will work with just two types of attitudes:

- **positive:** a positive attitude, feeling, evaluation, or opinion toward the object
- **negative:** a negative attitude, feeling, evaluation, or opinion toward the object

If a private state does not fit clearly into one of these categories, then please do not include a type for it. As our work progresses, the annotation of attitude type will be refined.

Following are some examples.

(HR6.2) And as usual, the portion [of the report] about China contains little truth and many absurdities, exaggerations and fabrications.

• **(writer) negative → report, expressed by subjective elements:** “contains little truth; absurdities, exaggerations, fabrications”

(HR6.3) “Its aim of the 2001 report is to tarnish China’s image and exert political pressure on the Chinese Government”, human rights experts said at a seminar held by the China Society for Study of Human Rights (CSSHR) on Friday.

• **(writer, human-rights-experts) negative → report, expressed by subjective element:** “tarnish”

• **(writer, human-rights-experts, report) negative → China; explicitly mentioned private state whose on is:** “aim”

(HR6.4) “The United States was slandering China again,” said Xirao-Nima, a professor of Tibetan history at the Central University for Nationalities.

• **(writer, Xirao-Nima) negative → United States; expressed by subjective element:** “slandering”

• **(writer, Xirao-Nima, United States) negative → China; explicitly mentioned speech event:** “slandering”

Comments: This example shows that nothing prevents the same text span from being annotated as two different things. The word “slanders” is a subjective element, with **(writer, Xirao-Nima)** as the source. It is also an explicit speech-event term, with **(writer, Xirao-Nima, United States)** as the source.⁵ That is, when we consider the level of **(writer, Xirao-Nima)**, the word “slanders” is a negative evaluation of the truthfulness of what the United States said. When we consider the level of **(writer, Xirao-Nima, United States)**, the word “slanders” communicates that, according to the writer, according to Xirao-Nima, the United States said something negative about China.

⁵**NOTE:** this speech event does not pass criterion C1 (see section 2.7.1) for being a significant-and-particular speech event – we are not told what the United States said. To date, our annotators have been restricted to identifying only significant-and-particular speech events, so they would **not** have annotated this speech event.

5 Sentence-level classes

Compatibility comment: As of August 2002, the accompanying instructions for performing the annotations do not refer to this attribute. It will be added in the near future.

This annotation is included to ensure that sentence-level judgments can be derived from the finer-grained judgments described so far in this document.

By default, we will define a **subjective sentence** to be any sentence that contains at least one **onlyfactive=no** tag or at least one **expressive subjective element** tag. Otherwise, the sentence is an **objective sentence**.

Since this rule is the default, you don't have to mark anything for sentences that it works for.

However, if the rule does not work for a sentence, you should add an annotation to the sentence, either **override: subjective sentence** or **override: objective sentence**. When might these be appropriate?

An **override: subjective sentence** is a sentence that you perceive as being opinionated even though it contains no expressive subjective elements or **onlyfactive=no** judgments. There is something else that makes it subjective, likely something contextual.

An **override: objective sentence** is appropriate for sentences with an explicit private state or speech event, or an expressive subjective element, which you see as being minor, and not significant enough for the sentence-level classification to be **subjective**.

6 Strength

The accompanying instructions for performing the annotations will specify exactly which strengths to indicate and how. Generally speaking, you will be asked to judge the strengths of private states designated by explicit **ons** as well as those expressed by subjective elements. In addition, you may be asked to judge the contribution particular words have (such as a word used as an **on**) to the strength of a private state.

7 Certainty

As the accompanying instructions for performing the annotations specify, you will be asked to specify your certainty in your annotations. The default value will be **certain**, so you only

have to add a tag if you are not certain.

Please feel free to indicate that you are not certain. With any natural language classification, there are borderline/unclear cases. Your judgments about your certainty will help us distinguish the clearer cases from the murkier ones.