Coding Manual for Distinguishing Subjective and Objective Sentences in Text: Draft 2

Janyce Wiebe Matthew Bell Julie Maples

March 4, 1999

1 Introduction

This document provides instructions for distinguishing "subjective" and "objective" sentences in text. The kinds of texts we will use are newspaper articles from the *Wall Street Journal*. The purpose of these instructions is to prepare the reader to mark sentences in news articles as being either subjective or objective. The general idea is to distinguish sentences presenting objective fact from those presenting opinions, evaluations, emotions, etc. An example of an objective sentence is:

```
<s><C  subjobj="o" certain=''0"> BELL INDUSTRIES Inc.
increased its quarterly to 10 cents from seven cents a
share. <\C><\s>
```

An example of a subjective sentence is:

```
<s><C  subjobj="s" certain=''0"> Ballooning at
the de rigueur hour of 6 a.m. held all the attraction for most people
of sunrise root-canal work. <\C><\s>
```

The first sentence deals with an undisputed matter in a straightforward, "factual" fashion, and the second relates the opinion of the masses towards ballooning.

Your mark-up to the text (henceforth referred to as "annotations" or "tags") will be used to develop a computer system to automatically distinguish between these two classes. The results of that system will be used to develop text processing computer applications.

We could mean many different specific things by "subjective" and "objective". The specific categories developed here are important for "document profiling" applications such as the following.

1. Consider searching for particular kinds of documents on the Web. Research tools are being developed in computer science to present the user with summaries and profiles of the documents, rather than just the title and first few lines, to help the user decide which documents to retrieve (*Information Retrieval*). A useful characteristic of the documents to include would be how opinionated it is.

2. Research tools are also being developed to assess how complex documents are, to determine how difficult to process they may be for computer systems, such as machine translation systems and information retrieval systems. Subjectivity is one important aspect of document complexity. For example, subjective documents often contain figurative language, which is very difficult for computers to process.

3. There are also a number of ways in which subjectivity is very important for developing computer systems to understand and produce language. We don't have room in these instructions to describe these in depth. Briefly, being able to determine which sentences are subjective can help systems:

   (a) determine what words mean

   (b) determine how the text is structured, and

   (c) in machine translation from, e.g., English to Spanish, produce a Spanish version that is similar in style to the English version.

There are other possible conceptions of the categories, which might support other applications. For example, another definition of the tagging task might be to find all of the facts in a document, even those communicated in evaluative sentences. The results of that task would support *information extraction* applications, in which facts are extracted from text. It is important to note that the tags assigned for that task would often be different from those assigned for the tagging task developed in this document. We hope to experiment with this "only the facts ma'am" alternative in the future. But for this work, please do not have it in mind.

The following sections will explain what we mean by "subjective" and "objective', and will provide you with some criteria and strategies to use in the annotation process.

2 The Basics

We now to turn to our definitions of "objective" and "subjective" sentences. **If the primary intention of a sentence is objective presentation of material that is factual to the reporter, the sentence is objective. Otherwise, the sentence is subjective.** In performing these annotations, the job is not to judge whether the material is actually true (after all, "objective reporting" often contains inaccuracies). Rather, the task is to judge whether it is being *presented* objectively as fact. In addition, note that the definitions do not preclude a subjective sentence from containing some factual information. After all, a single sentence can accomplish many different things.

On the other hand, an objective sentence need not be devoid of any hint of subjectivity. For example, the reporter may introduce some colorful wording to add interest, but retain a primary intention to objectively present factual material. For another example, informed projections or expectations, which often involve words like "likely" or "expect", can be objective. For example, "likely" does not make the following sentence subjective: "The computer paper is likely to run out by Wednesday; please order some now." In contrast, all things being equal, "likely" probably does make the following sentence subjective: "Smith is likely to win the election."

There isn't any formal, mathematical way to define categories such as subjective and objective. Generally speaking, what a sentence means comes down to how people understand it, in the context in which it appears. Please use your judgement, guided by the tagging criteria outlined in this document, and by the examples of tagged sentences provided to you.

There are just a few more points to cover in this section.

It is useful to identify the following two general classes of subjectivity.

1. The "Evaluative" Class, which contains
   - Evaluation
   - Opinion
   - Judgement
   - Emotion (or other internal, cognitive process)
2. The "Uncertain" Class, which contains:
   - Uncertainty
   - Doubt

- Speculation

When tagging, whose subjectivity do you consider? In a subjective sentence, subjectivity may originate with the reporter, with someone mentioned in the text, or with both. For example, "The music was beautiful" expresses the reporter's opinion and evaluation; "John thought the music was beautiful" expresses John's evaluation and opinion; and "John enjoyed the beautiful music" expresses both the reporter's and John's evaluations. When subjectivity of someone other than the reporter is expressed, then the sentence is either a "speech-event" or a "private-state" sentence (though not all speech-event and private-state sentences are subjective, as discussed in the next section). A "speech-event" sentence is about a speaking (or writing) event. A "private-state" sentence is about an internal state, such as believing, hoping, seeing, feeling, expecting, knowing, etc. The next section is about tagging speech-event and private-state sentences.

The final point to cover in this section is the format of the tags. Note that *compound sentences*, i.e., sentences with more than one main clause, are broken up into their *conjuncts* for this task. If this doesn't mean anything to you, don't worry. The sentences have already been broken up for you. Please just be sure to fill in all of the "subjobj" and "certain" attributes.

Every conjunct will have the symbol `<C>` before it and `<\C>` after. The `<C>` start tag will then have a series of attributes before the closing `>`. Possible values for *subjobj* are "s" for subjective and "o" for objective. In addition, you should place a numerical value from 0–3 in the *certain* attribute, **0 for most certain**, and **3 for least certain**. An example is shown below.

```
(C uid="26" subjobj="o" certain="0") Finding him became an obsession for
Mr. Stoll. </C>
```

Utterance 26 (i.e., uid="26") has been given the tag objective (subjobj="o") with certainty 0 (certain="0").

3 Speech-Event and Private State sentences

3.1 Introduction

Things get interesting when the sentence is about a speech event (e. g. , "He said ...") or private state (e. g. , "He believes that ...", "He hates ...", "He hopes ..."). Why? Because the sentence mentions the thoughts, beliefs, speech, etc. of an another person, other than the reporter. We have to consider the possible subjectivity of both.

An important problem we address in our research is determining the functions of speech-event and private-state sentences in the contexts in which they appear. Sometimes they should be treated as objective, sometimes they should be treated as subjective, and treating them appropriately is often important for understanding the article in which they appear. Subjective speech-event (and private-state) sentences are used to communicate the speaker's opinions, emotions, speculations, evaluations, etc. The reporter's main intention in these sentences is not to objectively present facts, but rather to present subjectivity of the speaker. The primary intention of objective speech-event (and private-state) sentences, on the other hand, is to objectively communicate material that is factual to the reporter. The speaker, in these cases, is being used as an authority, a reliable source of information. Reporters, especially in American newspapers, often present their facts in speech-event and private-state sentences in order to indicate their source of information.

Thus, we cannot simply treat all speech-event and private-state sentences as subjective or simply treat them all as objective.

Consider the speech-event sentence, "The suspect was apprehended at 5:15pm, Saturday June 15, the officer said. " (Assume that the officer's reliability is not in question.) A fact is being presented; the officer is the source of information, and the reporter isn't casting any doubt on his reliability. This sentence should be assigned the tag "objective". Similarly, the following private-state sentence should be tagged "objective" (again, assuming that the officers are reliable): "The officers discovered fingerprints on the gun which matched the fingerprints on the windowsill."

Contrast those sentences with the following (the first two are speech-event sentences and the second two are private-state sentences):

```
<s><C subjobj="s" uid=''20''> ''It may very well be what
the Japanese banks want," he told Radio New Zealand. </C></s>

<BR>(C subjobj="s" uid="50") Back in Berkeley, he was violently scolded
by a left-wing lady friend for consorting with such people. </C>

<BR>(C subjobj="s" uid="24") Mr. Stoll suspected the intruder was one of
those precocious students who has fun breaking into computers. </C>

<BR>(C subjobj="s" uid="49") He savored the humor of his uncombed appearance
among these buttoned-up chaps. </C>
```

These sentences are subjective. Their primary intentions are not to objectively present material that is factual to the reporter. Sentence 20 presents a speculation by the person referred to as “he”; sentence 50 describes a tongue-lashing “he” received from a lady; sentence 24 presents the suspicions of Mr. Stoll; and 49 communicates amusement on the part of “He”.

Now that you (hopefully) see the kind of distinction we want to make, let’s see how the tagging task is affected; you might be surprised.

You need to consider the entire sentence, not just the “outside level”. Consider sentences 20, 50, 24, and 49. One *could* argue that these sentences are objective; after all, it’s a fact, isn’t it, that the speech events did occur, and that the private-states were experienced (at least nothing suggests otherwise). In the case of 49, for example, there doesn’t seem to be any doubt that he did enjoy humor he found in the situation. But if we were to call these sentences objective, we would lose the distinction we were trying to make above.

Here is our strategy for tagging speech-event and private-state sentences. Work from the outside in. As soon as you find a reason to call the sentence subjective, stop; there is no reason to continue any further, because you already know the tag is “subjective”. But if you haven’t yet found a reason to call the sentence “subjective”, then continue, considering the next level of the sentence. An objective sentence is one where you get to the most nested level without finding a reason to call the sentence “subjective”. (In the spirit of considering main intentions only, consider the “main thing” at each level; please don’t use this paragraph as an excuse to call a sentence subjective at the merest hint of subjectivity.)

Let’s step through this process in a bit more detail, pointing out the types of things to consider as you step through a private-state or speech-event sentence.

At the outermost level, the sentence is about someone saying something or someone experiencing a private state. What could make the sentence subjective at this level? Doubt, evaluation, etc., expressed with words that are not in the scope of the speech-event or private-state term. For example (these examples are simple made-up examples to try to make the points clearer):

1. The *idiot* said that the computer is a Cray

2. “The computer is a Cray,” the *idiot* said.

3. John *probably* said that the computer is a Cray.

4. John *seemed* to know what his telephone number was.

The italicized words are not in the scope of the speech-event or private-state terms—*idiot* in (1) and (2) and *probably* in (3) are not part of what was said, and *seemed* is not part of what John knows. They make the sentences subjective (so you can stop now; you know the sentences are subjective).

Suppose that there isn't any subjectivity outside the scope of the private-state or speech-event term strong enough to make the sentence subjective. Now the thing to consider is the speech event or private state term itself. "said" is neutral, but consider "pontificated" and "blathered". These would make the sentence subjective, because they are evaluative toward the speaker. Speech-event terms like "scolded", "charged", and "argued" can make a sentence subjective, because they imply that the speaker is expressing his emotions, opinions, etc. Private-state terms as in "He became angry", "He hoped that...", and "He became obsessed..." can make a sentence subjective, because they introduce the emotions, desires, etc., of the experiencer. Private-state terms such as "believed" and "suspected" can make a sentence subjective, because they imply suspicions and opinions.

Suppose we haven't found anything yet to make the sentence subjective, so we need to continue. We now consider the object of the speech-event or private-state term (e.g., what was said, hoped for, hated, written, etc.). Of course, if there isn't an object (as in, e.g., "He spoke" or "He was angry"), we have already had to make our decision. Note that, in many cases, the lack of an object tends to make the sentence less subjective, because we are not being told the content of what was said, thought, etc. A sentence that merely alludes to the presence of a thought, for example, might be judged, in context, to be objective. However, whether this is possible depends on the term used. It is difficult to imagine an objective interpretation of "He was angry" (in which "angry" doesn't have an object), since "angry" so strongly evokes an emotion.

Anyway, suppose there is an object. Note that, if the object is itself about a private state or speech event, then we need to apply the same rules as used above (consider the words outside the scope of the private-state or speech-event term, and then the private-state or speech-event term itself). If it isn't, then we are at the innermost level of the sentence, and we finally have to decide whether or not the sentence is subjective. Ask yourself, is the object being objectively presented as material that is factual to the reporter? If so, the sentence is objective. Otherwise, the sentence is subjective. Compare the following sentences:

1. The suspect was apprehended at 5:15pm, Saturday June 15, the officer said (assuming the officer is reliable), versus

2. The AFL-CIO said the compromise has "adequate safeguards", so the youth are not exploited and older workers are not displaced.

Neither contains subjectivity outside the scope of the speech-event term, and in both, the speech-event term ("said") does not make the sentence subjective. The first is tagged objective, because what was said is presented as material that is factual to the reporter. The second is tagged subjective, because, in the context of the article the sentence appears in, the AFL-CIO is stating its opinion, which is not necessarily endorsed by the reporter.

4 Remember, There are Different Kinds of Subjectivity

We want to remind the reader that there are multiple types of subjectivity to consider, and not to get too focussed on any single one. Remember, the reporter must not only believe/endorse the material in a subjective sentence, the presentation must be primarily objective as well.

4.1 Negation

We have one final issue to discuss, that of how to tag a sentence if it deals with something that explicitly did not happen, was not said, was not thought etc. — in other words, what should be done in the case of *negation*.

All things being equal, the following uses of the italicized words are subjective:

- He *probably* said ...
- She *must* have been here ...
- He *would* have gone *if* ...

All things being equal, these words would probably warrant a "subjective" tag because of the degree of speculation involved. While it may seem that negation is similar, for our purposes we will treat it differently. If a sentence describes a state or event that has not occurred or will not occur, it is not necessarily disqualified from the "objective" class. For example, the following could easily be interpreted to be objective (depending on the context, of course):

- The stock market did not crash
- The stock market crashed

Thus, we see that while hedges such as "probably" suggest a sentence is subjective, negation itself often does not have the same effect.