

# Instructions for Tagging State-of-Affairs Types of Main Clauses in the Wall Street Journal

Julie Maples and Janyce Wiebe

## 1 Outline

The outline of this document is the following.

### Introduction

- What Makes a Clause a Main Clause

- Main Clause vs. Point of View Tagging

- Compound Sentences

### Clause Types

#### Speech Events

- Direct Speech

- Mixed Speech

- Other Speech Events

  - Indirect Speech

  - Sentences Whose Verbs Establish a Particular Focus

  - Sentences with ‘‘According To’’

#### Private States

- Descriptions of Various Cases

- The Role of the Experiencer

- Nonspecific Experiencers

- Partial List of Private-State Terms

#### ‘‘Other’’ Clauses

### Other Problem Issues

- Indicating Unusual Cases in the Tags

- Focus of Attention and Context Dependence

- Nested Compound Sentences

- Sentence Fragments

- First Person Speech

- Sentence Types Related to Subjectivity

  - Speculated States and Events

  - Negated States and Events

Suggested States and Events  
Mistakes in the Original Text or its Formatting  
Borderline and Ambiguous Cases

## 2 Introduction

This document describes hand-annotation (tagging) of main clauses of sentences in the Wall Street Journal. The tagging done in this task centers around the idea that sentences, consisting of one or more main clauses, group themselves into natural categories representing particular types of states and events. The project focuses on three general clause types: speech events, which describe any occurrences of communication; private states, which describe the internal conditions of a person, group, or other agent in the text; and clauses that do not fall into either of these two categories. Sentences or clauses about speech events are further subcategorized according to how the event is presented syntactically and how much of it is presented. A brief description of the clause types that have been tagged is given below; more in-depth descriptions are supplied at later points in the document.

The types of language used for speech events and private states are varied and productive (i.e., the reporter can combine different elements to produce a new way of presenting a private state or speech event). We cannot exhaustively describe all possibilities. What we hope we do in this document, however, is to give a sufficiently broad range of examples for the tagger to generalize to new situations encountered in the text.

The types of clauses we consider in this work are as follows (as you will see, “indirect speech” is not yet tagged as a separate category yet, but is lumped in with “other speech events”.):

### Clause Types

#### Speech Events

Direct Speech - the writer purports to present what the speaker actually said; such sentences contain quotation marks.

Indirect Speech - the writer paraphrases what was said; this is done without quotation marks and often uses the word “that” to introduce the paraphrase.

Mixed Speech - direct and indirect forms of speech are combined in a single sentence.

Other speech event clauses - speech events are expressed using various methods other than the preceding ones.

Private State - the internal state of an experiencer holding an attitude, optionally toward an object, is described.

Other - neither speech events nor private states are described.

The tags take the following form:

```
<MC soa="type">
```

where “MC” stands for “main clause”, “soa” for “state of affairs”, and “type” is one of five classifications.<sup>1</sup> The five possible “types” are “se.ds”, “se.ms”, “se.o”, “ps”, and “other,” described in the following subsections. Here is a sample main clause tag to illustrate its format:

```
<MC soa="se.ds">
```

You will be tagging data from an on-line version of the Wall Street Journal. In this data, the sentences have already been marked with the the sentence boundary markers

‘‘<s>’’ for the beginning and ‘‘</s>’’ for the end.

Paragraphs have already been marked similarly. The sentence and paragraph annotations do not concern this project except as a guide for placement of the main clause tags, which should immediately follow each beginning-of-sentence marker. As you will see in the data, a tag marking the end boundary of each main clause is placed immediately before the end-of-sentence marker, or right after the clause’s end if it is followed by another main clause.

## 2.1 The Classes

At this point in time, our categories of **speech events** are of different “grain sizes”. Two are very specific, and the other, “other speech events”, is a waste basket where we toss all of the other sentences about speech events. In the near future, we hope to break this last one down into subcategories.

Clauses containing speech are represented as one of the following:

```
se.ds    speech event in the form of direct speech
se.ms    speech event clauses that are mixtures of direct and
          indirect speech
se.o     clauses about speech events that contain indirect
          speech or other speech occurrences not described by
          either of the above categories
```

Because indirect speech was found to be closely related to several other types of speech events outside the realm of quoted speech, it has been included in the category “se.o” and tagged as such whenever it occurs.

The other tags are the following:

```
ps       private state
other    clauses that do not describe speech events or
          private states
```

---

<sup>1</sup>The term “state of affairs” is an academic term for a class that includes both events – something that happens, such as John hitting something or a volcano erupting—and states— such as something being six feet tall. Private states are types of states and speech events are types of events.

jm12-14: I'm leaving the following comment in case it describes any really important details that are still missing from this section. I think I added a few of them in the revisions I did on 11-22 though.

jm1021: Below are your comments from previous drafts, saying that we need more explanation of clauses here—a general definition, and stuff about the various types. I've tried to add some of the basics, but I don't think I can really do this without Quirk. Your previous comments are below.

jw8-6: we should say more about main clauses. We should mention the obvious point that a clause always contains a verb, so that, e.g., “She saw him sitting on the floor” contains two clauses, one embedded within the other, but the prepositional phrase “Into the scary dark night” is merely a phrase; it is not a clause (after all, it does not contain a verb).

jw8-6: say what a clause is, and give an example with a relative clause, a while clause, an object, etc. and point out the clauses, and intuitively describe why they are subordinate. Also, a conditional, etc. Most people don't know what a subordinate clause is.

## 2.2 What Makes a Clause a Main Clause

To understand the sentence types used in this task and how to choose among them, it is necessary to recognize exactly what we are calling the “main clause” of a sentence. The syntactic foundation of a main clause, in this task, consists of a sentence's main noun and verb groups. In fact, a main clause will always contain a noun and a verb, either explicitly or implicitly. At the same time, however, the “intuitive” approach of this task lets us use more than just a strict grammatical formula to determine the category to which a sentence belongs. In most cases, the first step toward classifying a sentence is simply to read it carefully while asking yourself the question, “What is the primary purpose of this sentence?” Often, first impressions are helpful in providing an answer. In more complex sentences, it may be helpful to imagine diagramming the sentence, paring off all subordinate clauses and other peripheral information until nothing is left but the most basic idea which the sentence is meant to communicate, and which is not subordinated syntactically by any other ideas. For sentences that fall into the “other” category, as well as for private states, this is often a simple process.

```
<s><MC soa="other"> Among 33 men who worked closely with the
substance, 28 have died -- more than three times the expected
number. </MC></s>
```

When all subordinate clauses are mentally removed, it is easy to see that the main clause of this sentence is “28 have died”, a statement that should be tagged “other”, since it is neither a speech event nor a private state.

```
<s><MC soa="ps"> For one thing, Pentagon officials, who asked
```

not to be identified, worry that the U.S. will have a much tougher time persuading Europeans to keep some short-range nuclear weapons on their soil once Soviet armored forces are thinned out.  
</MC></s>

Once again disregarding subordinate clauses, which include the mention of speech in this case (“asked not to be identified”), the main clause reads “Pentagon officials worry...”, which is a private state.

Classifying sentences that describe speech events is often more complicated because of their syntactic structure, which may make intuitions about the sentence’s purpose more difficult. In the most straightforward speech-event as well as private-state sentences, the main noun or noun phrase describes the agent, or the person, group or other entity in the text that is the speaker or experiencer. The main verb in such speech sentences is often “said”, “added”, or some other verb that presents the idea that an act of communication has taken place. The main clause of such a speech-event sentence is, in its simplest form, a phrase like “he said”. (In private states, the main verb may communicate the experience of the agent in a variety of ways, which will be discussed in a later section.) A phrase such as “he said” is the most obvious signal that a sentence or clause is a speech event. It is likely to be followed or preceded by a quote in a direct speech sentence, and in indirect speech by “that” and a paraphrase, as seen in the third example below. The word “that” may also be implied, as it is in the second example below. Note the comma in the first case; such a comma is common in direct speech sentences, when the “parenthetical” comes after the quoted part of the sentence.

<s><MC soa="se.ds"> "I was very frustrated," he said. </MC></s>

<s><MC soa="se.o"> The CIA told him it does not do domestic counterespionage. </MC></s>

<s><MC soa="se.o"> But the people familiar with the talks said that Warner isn’t expected to get any cash in the settlement. </MC></s>

Speech events are not always presented as simply as these, however. The manner in which the speech event is represented is more important in determining how it is classified than is the wording used to introduce it. No matter how complex a speech-event sentence may be, however, its primary purpose is to describe an act of communication. In sentences that introduce this act with a phrase such as “he said”, “she added”, or “officials responded”, the part describing the content of the speech event is the object of the clause, and is therefore not the part to be recognized in the sentence’s tag. The main clauses of the examples above, then, would be the following (the objects of the main clauses are enclosed in square brackets):

he said [I was frustrated]  
the CIA told him [it does not do domestic counterespionage]  
the people said that [Warner isn’t expected ...]

Understanding that the content of the speech event is really just the object of the sentence allows us to keep the focus on the speech event itself and the method with which the reporter presents it, which will determine how the sentence should be classified.

As mentioned above, the word “that” often plays an important role in sentences that describe speech. In particular, it appears either explicitly or implicitly in what is described below as “indirect speech”, a subcategory of se.o. It serves two purposes: to separate the content part of the speech event from the part mentioning the speech itself (i.e., “he said”), and to subordinate it syntactically from the rest of the sentence. This is achieved even when “that” does not actually appear in the sentence, as was shown in one of the examples above. Direct speech, category se.ds, does not have an implicit or explicit “that” between the main verb and the object which presents what was said. Contrast:

John said that the flowers are blue (indirect speech; se.o)

John said the flowers are blue (implicit ‘‘that’’; indirect speech; se.o)

‘‘The flowers are blue,’’ John said (direct speech; no implicit or explicit ‘‘that’’; se.ds)

In the case of mixed speech (se.ms), the sentence might be just like direct speech, except that some of what is said is not in quotes:

```
<s><MC soa="se.ms"> The changes were proposed in an effort to
streamline federal bureaucracy and boost compliance by the executives
"who are really calling the shots," said Brian Lane, special counsel
at the SEC's office of disclosure policy, which proposed the
changes. </MC></s>
```

This has the form of direct speech — there is no implicit “that” between the main verb and what was said.

In the other case of mixed speech (se.ms), the sentence has the form of indirect speech, but part of the object (the content of what was said) is enclosed in quotes. Examples are both parts of the following sentence:

```
<s><MC1 soa="se.ms"> The judge now says it was "unfortunate"
that he chose to write the letter </MC1><MC2 soa="se.ms"> but
says "there was certainly no intent to extort there." </MC2></s>
```

jw12-19: the second is se.ds – this is a mistake. The following example illustrates the se.ms case where the speaker and speech term are not included; here part of what is said is enclosed in quotes, and part isn't.

```
<s><MC soa="se.ms"> "Trying to answer that is suggesting all
kinds of theories," such as precisely where and how the brain
processes incoming signals from the eyes. </MC></s>
```

### 2.2.1 Main Clause Task vs. Point of View Task

It is very important for the tagger to keep in mind the distinction between the tagging task described here and the other task, *point of view* tagging, described in a separate document. The main clause tagging is concerned with the type of state of affairs the sentence is about. The other task is concerned with the *subjectivity* of the sentence, i.e., whether evaluation, judgement, belief, conjecture, etc., are expressed in the sentence, and if so, whose. Elements of the two tasks overlap in sentences that contain a subjective statement of some kind, and that describe the speech event or mental state of some agent in the text. The following example points out this often-subtle distinction.

```
<s><MC soa="other"> But the higher mortgage ceiling is  
only the starter kit for what Senator Alan Cranston and Majority  
Leader George Mitchell have in mind for housing. </MC></s>
```

**\*\*jw12-16!**: this is an out-of-place comment concerning what part of the sentence is being considered: one might wonder if this isn't a private state, since "being a starter kit" is pretty weak. But notice just how subordinated the private state term is. The ceiling is a starter kit for something which some people have in mind – the object is that thing, not the private state of having something in mind. If the sentence were, say, "The movie was a starter kit for new ideas", the sentence would be a private state. But this sentence is admittedly close to the border, because a liberal paraphrasing might describe the Senator's plans. **\*\***

This sentence contains a private state ("have in mind") as well as a subjective element showing the reporter's evaluation ("only"). Both of these will be considered in the point of view task, but for the purposes of this tagging task, the main clause alone ("the higher mortgage ceiling is only the starter kit") should be tagged "other." For further discussion and a more complex example of the interaction between the point of view and main clause tasks, see section (4.5.1).

Often a sentence that communicates a state or event of one type in its main clause contains information that, while not subjective, falls into one or more of the other clause categories. Here, too, the tagger should be careful to consider the type of the main clause only, since extra information appearing outside the main clause will be accounted for in the point-of-view task (so does not need to be accounted for in this task). Consider one of the examples used earlier:

```
<s><MC soa="ps"> For one thing, Pentagon officials, who asked not  
to be identified, worry that the U.S. will have a much tougher time  
persuading Europeans to keep some short-range nuclear weapons on  
their soil once Soviet armored forces are thinned out. </MC></s>
```

As was pointed out, the phrase "asked not to be identified" describes a speech event. The main clause, however, "Pentagon officials worry", describes a private state, and is what the sentence should be tagged for. The same principle should be used in tagging the following sentence:

```
<s><MC soa="se.o"> Describing itself as "asset rich," Sea Containers said it will move immediately to sell two ports, various ferries, ferry services, containers, and other investments. </MC></s>
```

Although this example contains a quoted phrase, “asset rich”, the main clause of the sentence is “Sea Containers said [it will move...]”, which is indirect speech or “se.o”. This sentence is an example of a case in which observation of syntax takes precedence over the intuitive approach. Many sentences that contain the quoted words of a speaker will have to be tagged as “se.o”, even though the inclination will be to call them “mixed speech”, because of their syntactic structure.

To avoid problems such as those mentioned above, the following three-part strategy for the tagging process is recommended, with each part having a particular focus:

1. In the first pass through the text, the tagger can classify and tag sentences according to first impressions.
2. The second time through should involve a closer and more careful look at the syntax of the sentence, as this is crucial to determining the type of the main clause and is often the source of confusion in this process. The tagger should specifically ask her/himself for each sentence whether it is the main clause only that has been tagged (or if more must be considered in accordance with the comments below in section 3.2.2).

It may help to simplify the sentence to something that is, for the purposes of this task, essentially the same, but is a simpler sentence. Consider this example:

```
<s><MC soa="other"> Free markets, free minds and free elections have an appeal that seems to get muddled only when delivered through U.N. organizations -- which of course are made up largely of governments that fear these principles at home. </MC></s>
```

This sentence is the same, for our purposes, as all of the following:

Free markets have an appeal that seems good.

Free markets have an appeal that is good.

Free markets have an appeal.

The phrase “have an appeal” in these sentences is a subjective element, and in the context the implicit experiencer is, by default, the reporter. This means that the phrase, which would in other contexts be a private-state term, is treated as a subjective element instead. So the sentence should be classified as “other”.

3. In the third pass, the tagger can clean up the tags, making decisions on the more difficult problem cases that are still outstanding.

### 2.2.2 Compound Sentences

Compound sentences—those having more than one main clause—occur frequently in the text. These should be given tags for each of their main clauses, and the “MC” tags should be given numbers, to aid in keeping track of the clauses (MC1, MC2, etc.). **The sentences have already been segmented into main clauses for you.** The end-of-clause tag for each clause should have the same number given in the beginning tag for that clause, such as “</MC1>”. Sentences have been found that have as many as six main clauses, and more are possible given the flexibility reporters are allowed and the effort they make to squeeze as much information as possible into every sentence of a news article. As the examples just below show, tags in compound sentences should be added before the first word of each clause, with a space separating the word and the tag. Main clauses may be separated by conjuncts, such as “and”, “or”, and “but”, and by marks of punctuation, such as semicolons, colons, dashes, and, rarely, commas. If the conjunct is a word (“and”, “or”, “but”), it is grouped with the clause that it introduces, and the tag for that clause should precede the conjunct. Marks of punctuation that separate clauses are grouped with the clause that they conclude, with the end tag following them. Here are some typical examples of compound sentences:

```
<s><MC1 soa="ps"> He believes in what he plays, </MC1>
<MC2 soa="other"> and he plays superbly. </MC2></s>
```

```
<s><MC1 soa="ps"> It's known for its cautiousness in part
because it has had little need for bold overseas ventures: </MC1>
<MC2 soa="other"> In the year ended March 31, 57.4% of its total
revenue came from office building management. </MC2></s>
```

```
<s><MC1 soa="other"> Rockefeller isn't the first overseas purchase
for Mitsubishi Estate -- </MC1><MC2 soa="other"> it has already
played a leading role in designing Los Angeles's Citicorp
Plaza. </MC2></s>
```

```
<s><MC1 soa="se.o"> Mr. Nahas, who single-handedly provoked a
one-day closure of Brazil's stock markets in June when he failed to
honor a debt of $31.1 million owed to his brokers, yesterday blamed
his predicament on the president of the Sao Paulo stock exchange;
</MC1><MC2 soa="other"> a few days before Mr. Nahas's failure, the
exchange raised the required margin on stock-margin
transactions. </MC2></s>
```

As is shown above, the clauses of a compound sentence can be classified as different types. If they are of the same type, however, both should be tagged as that type as if they had appeared separately (see for example the third sentence above, in which both main clauses are tagged “other”).

Sentences classified as compounds also include those containing multiple main verbs, for some of which the subject (“he” in both of the following examples) is implied:

```
<s><MC1 soa="other"> He gave up seven hits, </MC1>
<MC2 soa="other"> walked five </MC2><MC3 soa="other"> and didn't
get a decision. </MC3></s>
```

```
<s><MC1 soa="other"> Sometimes, if the hacker seemed to be into a
sensitive file, he would drag his keychain across the terminal to
create static </MC1><MC2 soa="other"> or slow the system down to
frustrate his quarry. </MC2></s>
```

Less common are compound sentences in which all or part of the verb is implied in some clauses. In the first sentence below, the ellipsis occurs with the auxiliary verb, or the part of the verb whose purpose is to help form the tense and aspect of the main verb. In the second sentence, the main verb is implied in the second and third clauses.

```
<s><MC1 soa="other"> So far this year, the index of leading
indicators has risen in four months, </MC1><MC2 soa="other"> fallen
in four months </MC2><MC3 soa="other"> and remained unchanged in
the other month. </MC3></s>
```

```
<s><MC1 soa="other"> Delivery of the first aircraft is set for
early November, </MC1><MC2 soa="other"> a second for December
</MC2><MC3 soa="other"> and two for April 1990. </MC3></s>
```

And even more unusual, but still taggable as a compound sentence, is the “bisected” form, in which one clause interrupts another clause that is already in progress:

```
<s><MC1 soa="other"> Still, the restaurant's ever-changing menu of
five-course dinners -- <MC2 soa="other"> it supposedly hasn't
repeated a meal since opening in 1971 <\MC2> -- requires constant
improvisation. <\MC1></s>
```

```
<s><MC1 soa="ps"> The general public may not know his
name, <\MC1> <MC2 soa="other"> but he's famous -- <MC3
soa="other"> make that infamous <\MC3> -- in advertising circles:
<\MC2> <MC4 soa="other"> A marketing consultant, he pioneered
slashing ad agency commissions, to the delight of advertising
clients and the dismay of agencies. <\MC4></s>
```

Note especially the arrangement of the tags in sentences such as the above: their placement at the beginning and end of each clause allows the tagger to recognize the clauses’ boundaries even if those boundaries overlap. jw12-16: please leave a blank line for paragraph breaks. Latex doesn’t see these indentations.

While the conjuncts and punctuation marks mentioned above are good indicators of compound sentences, they should be observed carefully, because many sentences contain these elements but are not compounds. In particular, sentences that begin with “and”, “or” or “but” and consist of only one main clause are to be tagged as regular independent clauses. In such cases, “and”, “or”, and “but” perform the same syntactic function as “However,” or “In addition,” when these are used at the beginning of a sentence. Here are some examples.

```
<s><MC soa="ps"> But you can't dismiss Mr. Stoltzman's  
music or his motives as merely commercial and lightweight. </MC></s>
```

```
<s><MC soa="ps"> And their suspicions of each other run deep.  
</MC></s>
```

Many non-compound sentences can also be found which use any of the conjuncts or punctuation marks listed above to separate subordinate clauses rather than main clauses:

```
<s><MC soa="se.o" ws"> Many of the letters maintain that  
investor confidence has been so shaken by the 1987 stock market  
crash -- and the markets already so stacked against the little guy  
-- that any decrease in information on insider-trading patterns  
might prompt individuals to get out of stocks altogether. </MC></s>
```

```
<s><MC soa="se.o" ws, sc"> The filing also said the trust  
is considering a sale of Manville securities in the open market; an  
extraordinary dividend on the common stock; or a recapitalization of  
Manville. </MC></s>
```

To be tagged as a compound sentence, all of the main clauses of a sentence must stand on their own, either because they contain main noun and verb groups or because they have the main noun or verb implied by another independent clause.

A sentence that has multiple main clauses which are objects of the same parenthetical should not be divided and tagged as a compound sentence, since the parenthetical itself is considered the main clause. This issue will receive further discussion at a later point, but an example is shown here for its relevance to this section.

```
<s><MC soa="se.ds" fp"> "You bet attention," I  
yelled back, leaping atop the propane tanks, "I'm wearing alligator  
loafers]" </MC></s>
```

### 3 Clause Types

The following sections describe the five different main clause types: direct speech, mixed speech, other speech events, private states, and other clause types not included under one of these headings. Each section contains tagged examples of sentences of these types, and

describes some of the problems that each type presents. All tagged examples are actual sentences from the Wall Street Journal text that has been tagged. After the descriptions of the main clause types, a discussion of some kinds of sentences which present variations on or deviations from these types will be given.

### 3.1 Speech Events

An important question in specifying speech event categories is whether the reader's perception of the category is based on event type, syntactic structure, reporter intention, or other factors. While many of the same questions and solutions apply to all the categories of speech events, the nature of the the direct speech classification makes this category a smaller, more specific one, and so allows it to avoid many of the questions and exceptions that arise for the other categories. Because of its close relationship to direct speech, the mixed speech category is also a fairly small one. Overall, though, there is a great deal of variety in how speech events are presented in text. Because of this wide range of speech occurrences, especially in the area of paraphrased or described speech, the category "other speech events" was created to encompass indirect speech and any related speech types that do not correspond to the direct or mixed speech categories. The tag "se.o" represents all occurrences of indirect speech, then, as well as additional speech event types described in section (3.1.3).

The distinction among "se.ds", "se.ms" and "se.o" concerns, among other things, the object of the speech event—that is, the section of the sentence set apart by quotation marks or subordinated to a construction like "he said that". (For private states, we are **not** making a corresponding distinction among methods of presentation. There is only one private state category.) The following example shows a sentence that seems to present several different states of affairs, all of which are subordinated to the parenthetical at the end.

```
<s><MC soa="se.o"> Northwest Airlines settled the
remaining lawsuits filed on behalf of 156 people killed in a 1987
crash, but claims against the jetliner's maker are being pursued, a
federal judge said. </MC></s>
```

The above example shows the importance of reading sentences all the way through before deciding on their type, as the element that determines the choice may not appear until the very end.

In this task, no category distinction is made between spoken and written words, since similar words and syntactic structures are used to present both writing and speaking in text. However, as a practical distinction, the written occurrences are given the annotation "ws" in the comment field of the tag. Direct, indirect and mixed speech may all occur in written form, as shown below.

```
<s><MC soa="se.ds"> "The SEC has historically paid
obeisance to the ideal of a level playing field," wrote Clyde
S. McGregor of Winnetka, Ill., in one of the 92 letters the agency
has received since the changes were proposed Aug. 17. </MC></s>
```

<s><MC soa="se.ds"> "Apparently the commission did not really believe in this ideal." </MC></s>

<s><MC soa="se.o" nh"> Many of the letters maintain that investor confidence has been so shaken by the 1987 stock market crash -- and the markets already so stacked against the little guy -- that any decrease in information on insider-trading patterns might prompt individuals to get out of stocks altogether. </MC></s>

<s><MC soa="se.ms" nh"> The Committee on Federal Regulation of Securities for the American Bar Association argues, for example, in its lengthy letter to the SEC, that the proposed changes "would substantially improve the {law} by conforming it more closely to contemporary business realities." </MC></s>

Sentences whose verbs are less obvious signals of speech events may also describe written speech:

<s><MC soa="se.o"> In "The Cuckoo's Egg" (Doubleday, 326 pages, \$19.95), he spins a remarkable tale of his efforts over 18 months to catch a computer spy. </MC></s>

<s><MC soa="se.o"> In an Oct. 19 review of "The Misanthrope" at Chicago's Goodman Theatre ("Revitalized Classics Take the Stage in Windy City," Leisure & Arts), the role of Celimene, played by Kim Cattrall, was mistakenly attributed to Christina Haag. </MC></s>

Written speech events may take a variety of forms depending on their source as well as the reporter's use of this source in the structure of the article. Sources of written speech include books, tests, laws, rulings, research findings, written statements, company policies, letters, statistical figures, reports, and treaties. The reporter may treat these sources as if they were human agents, or the human behind the written source may be recognized. In either case, the communication involved is tagged as a speech event, and the additional annotation "nh" is used to mark cases in which a non-human agent, with no recognized human behind it, is responsible for the speech event. Written speech and non-human agents are often, though not always, found together.

<s><MC soa="se.o" nh"> French consumer prices rose 0.2% in September from the previous month and were up 3.4% from a year earlier, according to definitive figures from the National Statistics Institute. </MC></s>

While it is helpful to distinguish between human and non-human agents wherever possible, this distinction is not a significant goal of this tagging task and should not be given precedence over the decisions concerning main clause types.

Not all communicative acts in a written form will be categorized as speech events; below is an example whose verb, “publish”, probably involves words, but puts the focus on the action performed rather than on communication.

```
<s><MC soa="other"> Companies would be compelled to publish  
in annual proxy statements the names of insiders who fail to file  
reports on time. </MC></s>
```

jw12-19: this “publish” is like “report”; this should be se.o. For more about deciding if a sentence focuses on a speech event or an action, see section 4.2.1.

### 3.1.1 Direct Speech

This category, like the other speech event categories, is based on an understanding of what the reporter is doing with the speech event of another agent. At any time that the reporter purports to preserve an agent’s actual words, rather than paraphrasing or describing them, the clause or sentence is to be classified as direct speech. Quotation marks are the only real guarantee that a speech event is being preserved word-for-word as it occurred; therefore, direct speech sentences will always contain speech events that are enclosed in quotation marks. Direct speech sentences may or may not include a parenthetical such as the following:

```
“...”, he said  
“...”, she told reporters.
```

Note that both sentences below are labelled direct speech (“se.ds”) because both are the quoted words of the speaker. The first but not the second contains a parenthetical.

```
<s><MC soa="se.ds"> "You can't give it up that easily," he  
says. </MC></s>  
<s><MC soa="se.ds"> "I tried." </MC></s>
```

Remembering that the parenthetical may be implied is important to an understanding of what constitutes the main clause when it comes to direct speech sentences: the speech event, rather than its contents.

Above it was stated that what is said in a direct-speech sentence must be enclosed in quotes. In a series of quoted statements, one or more of the quotation marks might not appear in the sentence itself, but might appear in earlier and/or later sentences. That is, a series of sentences might all be covered by one set of quotes. The following passage is an example. The opening quotes are in the first sentence and the closing quotes are in the third one, so that the middle sentence is covered by the quotes in the earlier and later sentences.

```
<s><MC soa="se.ms"> Instead, this official said, "This is vintage  
George Bush. </MC></s>  
<s><MC soa="se.ds"> This was George Bush's own idea. </MC></s>  
<s><MC soa="se.ds"> It's George Bush wanting to meet a foreign  
leader and talk to him directly." </MC></s>
```

All three sentences should be tagged as direct speech, even though none of them contains a full set of quotation marks, because they are all enclosed by the quotation marks that begin and end the quoted statement, and because all attempt to capture the original words of the speaker.

As in every category, there are exceptions which seem to contradict the rules for this clause type. Note the following, which is direct speech but uses the phrase “said that”, usually exclusive to indirect speech, to introduce the quote:

```
<s><MC soa="se.ds"> Mr. Johnson also said that "in the scheme
of things, these things are minor." </MC></s>
```

Although it looks like an indirect speech event, this example is tagged as direct speech because it presents the agent’s statement as being preserved word-for-word as it occurred. As was stated before, the manner in which the speech event is represented, rather than the wording used to introduce it, is the standard that should be used to determine the type of the clause.

### 3.1.2 Mixtures of Direct and Indirect Speech

The category of mixed speech (“se.ms”) is composed of clauses that contain a combination of direct (quoted) and indirect (paraphrased) speech. A mixed-speech sentence often seems to be the reporter’s attempt to control the tone of an article, by allowing agents to speak where their words have the most effect, and rephrasing or passing over other words that are not as important for meeting the goals of the article. At other times, it appears to be a way for the reporter to remain objective, or perhaps to keep an appearance of objectivity, while communicating a certain view. By inserting quoted speech where a statement or opinion is needed, the reporter can let the agent do the talking instead of making what would otherwise be seen as a personal comment on an issue, or he or she can use the agent’s words to make a personal comment. Some examples of sentences to be tagged “se.ms” are the following.

```
<s><MC soa="se.ms"> It was in Jaguar’s best interests "for the
company’s future to be assured and the present climate of
uncertainty resolved as quickly as possible," Mr. Ridley
said. </MC></s>
```

```
<s><MC soa="se.ms"> The Teachers Insurance and Annuity
Association-College Retirement Equities Fund said it will introduce
a stock and bond fund that will invest in "socially responsible"
companies, and a bond fund. </MC></s>
```

```
<s><MC soa="se.ms"> The company said the plan, under
review for some time, will protect shareholders against "abusive
takeover tactics." </MC></s>
```

Often mixed speech finds a more practical purpose, when some background is needed for a statement made by an agent in the article. The examples below show the use of mixed speech

by the reporter to provide information that the reader might need in order to understand or correctly interpret the agent's statement.

```
<s><MC soa="se.ms"> Because of their longstanding rivalry,
GM just "wants to make sure Ford pays a huge packet for
(Jaguar)," said John Lawson, an auto analyst at London's
Nomura Research Institute. </MC></s>
```

```
<s><MC soa="se.ms"> The AFL-CIO, which previously
opposed the administration's subminimum idea, said the
compromise has "adequate safeguards, so the youth are not
exploited and older workers are not displaced." </MC></s>
```

Karen: new, more specific definition of mixed speech jw12-16!: I just expanded this to include examples like the one two above – the one with “longstanding rivalry” – se.ms sentences that have the syntactic form of direct speech.

Syntactically, mixed speech is either the same as direct speech, with the difference being that not all of the object is contained in quotes (see the example two above, beginning with “Because”), or is a combination of indirect speech (**not** other types of “se.o”) and direct speech. Let's consider the second case. Specifically, this consists of direct speech nested within an indirect speech clause. The main idea to use when looking for this kind of mixed speech clauses is that the quoted part of the speech event in mixed speech must be subordinated to the part introducing the speech event. Many sentences contain both quoted and unquoted speech, but a clause is only to be tagged as “se.ms” if the quoted part remains “inside” the dominant speech event. jm12-14: Is the preceding sentence still true at this point, as far as the goals of the task are concerned? jw12-16: I changed it some.

Often the reporter will use a word or words that describe speech, such as “insisting” in the example below, to introduce a quote in the middle of a sentence, and in so doing he or she takes the reader out of the indirect speech event, and so removes the option of a “mixed” construction. Any representation of speech that follows such a word (“nuances”, below) is no longer part of the mixed speech event, because it is at that point no longer part of the first clause. jw12-16: I think the end of the previous sentence was wrong; I changed it.

```
<s><MC soa="se.o"> Mr. Mulford said reports of tension
between the Treasury and Fed have been exaggerated, insisting
that they involved "nuances." </MC></s>
```

jw12-16: I added the following. In other words, we need to distinguish two clauses in this sentence: The “Mr. Mulford said” clause and the “insisting that they involve...” clause. The first is the main clause, and the second is a subordinate clause. But the second is NOT part of the first clause – the second clause is not part of the object of the main clause. The “insisting” starts things over, in a sense, beginning a new speech event.

The second clause of the following example illustrates the same practice on the part of the reporter. The addition of “he said” before the quoted part (“less harsh”) causes a departure

from the speech event described by the clause's main verb "cited". Once these two speech events have been separated, the clause can no longer be tagged as mixed speech, but must be called "se.o".

```
<s><MC1 soa="se.ms"> Mr. Kennedy suggested Democrats
"yielded a great deal" on the size of the increase, </MC1>
<MC2 soa="se.o"> but he cited concessions from the White
House on the training wage, which he said make it "less harsh."
</MC2></s>
```

jm12-14: the preceding examples and discussion are still correct, aren't they? I still have the sentences tagged "se.o" in my file. jw12-16: yes, these are correct (albeit strange).

In the next example, the reporter has inserted quoted words ("very well") without taking the reader out of the agent's speech, making the sentence a mixed speech event. It should be noted that as mixed speech contains some amount of paraphrase, the reporter is involved in the whole sentence to some degree, but the words that prevent a speech sentence from being tagged "se.ms" are the ones which could belong only to the reporter, such as "insisting" in the previous example, and which could not possibly be attributed to an agent speaking within the text. In the next example, the only words that must be attributed to the reporter are "Moody's said". There are no interrupting words that prevent the sentence from being tagged as mixed speech.

```
<s><MC soa="se.ms"> Moody's said Houston
Lighting's current situation has some positive aspects, including
managing "very well" the construction and commercial operation
risks of Units 1 and 2 of the South Texas Project nuclear power
plant. </MC></s>
```

Following are more sentences that were tagged "se.o" instead of "se.ms" because the quoted words they contain are outside of the speech event in the main clause.

```
<s><MC soa="se.o"> The judge, suspended from his bench
pending his trial, which began this week, vehemently denies all the
allegations against him, calling them "ludicrous" and "imaginative,
political demagoguery." </MC></s>
```

jw12-16! Please check the following. Julie agreed to flipping around our treatment of these sentences. Ok Karen? Note that sentences that introduce speech events using the phrase "according to" and that contain part of the object in quotes **will** be tagged as mixed speech. Admittedly, technically they do not fit into either of the base types of se.ms:

1. "according to" cannot be used with direct speech without sounding strange. For example:  
"Interest rates will rise," according to an industry source.

2. And sentences with “according to” are not classified as indirect For further discussion of “according to”, see section 3.1.3.

But such sentences are so similar to other se.ms sentences, we opted to allow this possibility. Thus, the following are all tagged se.ms:

```
<s><MC soa="se.ms" ws, change in data">  
"My name is judge," Judge O'Kicki  
told a car salesman in Ebensburg when he bought a new red Pontiac  
Sunbird in October 1984, according to the grand-jury  
report. </MC></s>
```

```
<s><MC soa="se.ms" ws, change in data">  
When the judge bought his new  
Sunbird from James E. Black Pontiac-Cadillac in Ebensburg five years  
ago, the dealership had "certain apprehensions" about the judge's  
reputation, according to the grand-jury report. </MC></s>
```

```
<s><MC soa="se.ms" ws, change in data">  
"There was no loan, there is no loan,  
there never shall be a loan," the judge wrote the bank on his judicial  
stationery, according to the report. </MC></s>
```

The following, which might be seen as similar to the above sentences, is also tagged se.ms:

```
<s><MC soa="se.ms" in data''>  
The total relationship, as Mr. Lee sees  
it, is "the flow of dollars to the U.S. to fund the deficits, the  
investments the Japanese are making in the U.S. in order to satisfy  
American demand that American products consumed in America  
should be made as much as possible in America by Americans with  
Japanese technology and capital." </MC></s>
```

Some examples of speech events feature only one or two words of the original speaker quoted directly. Because they contain both quoted and paraphrased speech, these are to be tagged as mixed speech events no matter how many words are placed in quotation marks, as long as these words are part of the sentence's main clause rather than a separate subordinated clause. jw12-16: I added “separate”

```
<s><MC soa="se.ms"> Expect "tricky" stuff from pitchers, says  
Mr. Weaver, the manager. </MC></s>
```

```
<s><MC soa="se.ms"> Jerome J. Jahn, executive vice president and  
chief financial officer, said Mr. Rubendall was resigning by "mutual  
agreement" with the board. </MC></s>
```

The same is true for sentences in which only one or two words attributed to the speaker are excluded from the quote (“they” in this example):

```
<s><MC soa="se.ms"> In a statement, Jaguar's board said
they "were not consulted about the (Ridley decision) in advance and
were surprised at the action taken." </MC></s>
```

jw12-16: Note that if “they” were included within the quotes in this last sentence, the sentence would be direct speech (see the example at the very end of section 3.1.1).

Most difficult of all are sentences that feature two types of nested information: quoted speech is contained within an indirect speech event, and the speech event of one agent is related as a part of the speech of another agent.

```
<s><MC soa="se.ms"> "Don't you think I ought to get a
commission . . . or part of your fee in this case?" Mr. Green said
the judge asked him. </MC></s>
```

Careful examination of this sentence’s structure shows that it is an indirect speech event at its highest level (“Mr. Green said”, with an implicit subordinator “that”) and that it contains a direct quote (“Don’t you think...”) which is part of that speech event. jw12-16: Note that the sentence is not se.ds, because there is part of what Mr. Green said that is NOT in quotes — “the judge asked him” paraphrases something that Mr. Green said! Even though the quoted speech is presented as originally being the words of another agent, it is also part of the statement by Mr. Green which is the focus of this sentence. Syntactically, the entire sentence is subordinated to the words “Mr. Green said”, and can therefore be tagged “se.ms”.

The following sentence, also a difficult case, was tagged “se.ms” because its main clause (“they are still wrestling over... questions”) includes a quoted example of the “questions” that it describes (“What is a tank?”). Because this quote was seen as an integral part of the speech event, and because it is subordinated to it syntactically, this sentence is also classified as “se.ms”.

```
<s><MC soa="se.ms"> But even though NATO negotiators have
only 10 months left under the Bush timetable, they are still
wrestling over such seemingly fundamental questions as "What is
a tank?" </MC></s>
```

One important detail to note in identifying sentences with mixed speech is the difference between quoted speech and words that are placed in quotation marks for other reasons. The following sentence contains two phrases enclosed in quotation marks which are not speech events, but rather the titles of musical numbers. Such cases should not be treated in the same way as clauses with speech events that appear in quotes, but should be tagged “other” (or whatever class is appropriate based on the main clause).

```
<s><MC soa="other"> An improvisational section was built
around pieces by Mr. Douglas, beginning with "Golden Rain," a lilting,
```

laid-back lead in to the uptempo "Sky," which gave Mr. Stoltzman the opportunity to wail in a high register and show off his fleet fingers. </MC></s>

As can be seen in this section, examination of syntactic structure is more important in classifying mixed-speech sentences than in identifying most other sentence types, for which the more intuitive approach could be used.

### 3.1.3 Other Speech Events

With most of the other kinds of sentences about speech events, the reporter's purpose is to paraphrase, acknowledge, evoke, describe, or refer to a speech event rather than to repeat it or to present its contents. As was stated earlier, indirect speech, in which the reporter attempts to paraphrase another agent's speech event, is included in this category. Also included in this category are sentences that make reference to events involving speech or writing, but do not attempt to capture the contents of the actual words. These kinds of sentences, which in this task are placed in the category "se.o", are among the most difficult to tag for two main reasons: (1) they are at the "edge" of the speech-event category, coming right up against the border of the "other" category, and (2) they are often vague. Not all sentences containing words that could describe speech events should be classified as "se.o"; judgment calls will be necessary concerning how focal the speech event is in the sentence and its context. These judgements will be difficult in some cases. It is often difficult to decide whether the line has been crossed from the speech-event category into the "other" category. Sentences in the "se.o" category provide perhaps the greatest opportunity, in the flexibility with which they can be interpreted, for the reporter to shape the reader's perception of both the form and the content of the speech event being described. Following are some examples of sentences tagged "se.o".

<s><MC soa="se.o"> Mr. Stoltzman introduced his colleagues: Bill Douglas, pianist/bassoonist/composer and an old buddy from Yale, and jazz bassist Eddie Gomez. </MC></s>

<s><MC soa="se.o"> New Zealand's finance minister, David Caygill, lashed out at such suggestions. </MC></s>

<s><MC soa="se.o"> When he sent letters offering 1,250 retired major leaguers the chance of another season, 730 responded. </MC></s>

<s><MC soa="se.o"> Mr. Ridley hinted at this motive in answering questions from members of Parliament after his announcement. </MC></s>

<s><MC soa="se.o"> Germany's trade surplus is largely with other

European countries rather than with the U.S., Mr. Mulford acknowledged. </MC></s>

<s><MC soa="se.o"> A SEC proposal to ease reporting requirements for some company executives would undermine the usefulness of information on insider trades as a stock-picking tool, individual investors and professional money managers contend. </MC></s>

As mentioned above in section 3.1.1, the direct speech category is deliberately being kept a small, tight category; if a sentence is similar to direct speech but deviates from that class in any significant way, then it should probably be classified as “se.o”. For example, the following have the prototypical syntactic form of a direct speech sentence, but do not contain quotation marks. The omission of the quotation marks suggests that there is some amount of paraphrase being performed.

<s><MC soa="se.o"> No matter who owns PS of New Hampshire, after it emerges from bankruptcy proceedings its rates will be among the highest in the nation, he said. </MC></s>

<s><MC soa="se.o"> Intervention, he added, is useful only to smooth disorderly markets, not to fundamentally influence the dollar’s value. </MC></s>

For this reason, these sentences should be classified as “se.o” rather than “se.ds”.

## Indirect Speech

Indirect speech is the result of the reporter’s attempt to present a speech event by paraphrasing what was said, rather than purporting to present what was actually said. For a sentence to be understood as a prototypical indirect-speech sentence, the syntactic options are limited: the sentence is a complex sentence<sup>2</sup> with the paraphrase of what was said as the object of the sentence. That is, the sentence prototypically has the form ⟨agent⟩, ⟨speech-verb⟩, “that” ⟨paraphrase⟩ (but the subordinator “that” is often implicit). For example:

<s><MC soa="se.o"> But the people familiar with the talks said that Warner isn’t expected to get any cash in the settlement. </MC></s>

<s><MC soa="se.o"> The other 40% said the Fed ought to worry less about inflation and bring interest rates down. </MC></s>

The speech-event term may be something less direct than “said”:

---

<sup>2</sup>A sentence with a subordinated clause as object

<s><MC soa="se.o" nh"> Many of the letters maintain that investor confidence has been so shaken by the 1987 stock market crash -- and the markets already so stacked against the little guy -- that any decrease in information on insider-trading patterns might prompt individuals to get out of stocks altogether. </MC></s>

<s><MC soa="se.o" ws"> But internal church documents clearly show that at its publicity-seeking heights, as when it organized a spectacular Yankee Stadium bicentennial rally in 1976, there actually were only about 2,000 full-time Unification Church members in the U.S. </MC></s>

### Sentences Whose Verbs Establish a Particular Focus

At times the reporter may attempt to add a particular focus or slant to a speech event, or even to inject some personal comment on the issue involved, without quoting directly any part of the speech event itself. To communicate the occurrence of speech in such cases, he or she is likely to use a verb other than "said" in order to establish the desired focus, such as "defended", "hurled", "accused", "lashed out", and "complicated" in the examples below. Sentences using such verbs, while they are tagged as "se.o", deviate from the indirect speech category in that their main clauses typically present relatively less of the content of what was said than a prototypical indirect speech sentence would. In other words, the sentence presupposes the speech event, or makes the reader aware that a speech event has occurred, but the purpose of the sentence, at least at its most fundamental level, is not to paraphrase or present **what** was said. jm12-14: Is this the kind of sentence that will go in the "non-neutral speech verb" subcategory later? If so, should there be a note here to that effect, and should this section talk about verbs like "concluded"? jw12-16: I don't think we need to describe this for this study. The idea is that we later plan, in a new tagging task, to subcategorize the speech event categories to distinguish neutral speech terms like "said" from less neutral terms like those above. Another example is "concluded": this can be used as either a private state or speech term, depending on the context. But even when used as a speech term, it communicates more than that the person said something. For this tagging task, however, these distinctions will not be made.

<s><MC soa="se.o"> Treasury Undersecretary David Mulford defended the Treasury's efforts this fall to drive down the value of the dollar, saying it helped minimize damage from the 190-point drop in the stock market Oct. 13. </MC></s>

<s><MC1 soa="se.o"> The two sides in the legal battle have hurled accusations of duplicity at each other for weeks, </MC1>  
<MC2 soa="se.o"> and both Warner and Sony have accused each other of trying to sabotage each other's prospects for success in

the entertainment business. </MC2></s>

<s><MC soa="se.o"> New Zealand's finance minister, David Caygill, lashed out at such suggestions. </MC></s>

<s><MC soa="se.o"> The Soviets complicated the issue by offering to include light tanks, which are as light as 10 tons. </MC></s>

The use of intuition about the reporter's purpose, and how it is reflected in his or her choice of words, is especially important in identifying sentences of this kind.

### Sentences with "According To"

A particularly difficult type of speech event sentence, and one found frequently in the text, involves the use of the phrase “, according to ...”. The difficulty in tagging sentences containing this phrase comes from the fact that, although it appears in the syntactic role of a parenthetical, a phrase such as “according to company officials” is not a main clause, like the parenthetical “,he said”, because it contains no verb. Such a sentence seems to be an example of the type of subjectivity mentioned above: the use of “according to” avoids committing the reporter to any precise interpretation or treatment of the speech act, but instead provides syntactic and subjective freedom in the structure of that particular sentence, enabling the reporter to easily connect it to the surrounding discourse (note the subjective element (“not surprisingly”) in the first example below, and the liberties taken with the testimony (“extorted”, “muscled”, and “bullied”) in the second example's verbs). Because of this freedom, however, sentences with “according to” are not included in the indirect speech category, which attempts a more structured representation of speech events, but only in the broader category of “other speech events”. The fact that we are considering the phrase “according to” as a signal of speech event sentences at all is an inconsistency in the tagging task, as the task is supposed to target main clauses only, but this exception is being made to be consistent in spirit with the definition of the “se.o” category: this type of sentence suggests a paraphrase, and, through the use of “according to”, brings another speaker besides the reporter into the sentence. Thus, even though it is inconsistent with the goal of tagging only the main clause, sentences like the following should be tagged “se.o”.

<s><MC soa="se.o"> Campbell Soup, not surprisingly, doesn't have any plans to advertise in the magazine, according to its spokesman. </MC></s>

<s><MC soa="se.o" nh, ws"> According to testimony in a public, 80-page grand-jury report handed up to the state attorney general, Judge O'Kicki extorted cash from lawyers, muscled favorable loans from banks and bullied local businesses for more than a decade. </MC></s>

## 3.2 Private States

*Private states* are states of an experiencer holding an attitude, optionally toward an object. Varieties of private states include intellectual ones, such as someone believing, wondering, or knowing something; emotive ones, such as someone hating something or being afraid; and perceptual ones, such as someone seeing or hearing something. Thus, private states are states that are not open to objective observation or verification (Quirk et al. 1985). That private states are presented at all in text is the result of the reporter's freedom to express and sometimes predict the views and experiences of the agents that he or she is writing about. The instructions for the point-of-view tagging task contain more details about this clause type.

### 3.2.1 Descriptions of Various Cases

Because of the variety and complexity found in the occurrences of private-state sentences in the text, the most effective way to describe this category will be the use of a representative group of actual and prototypical sentences. Following are some examples from the tagged text of sentences that describe private states.

```
<s><MC soa="ps"> Some of the game's reigning philosophers  
dislike the idea of middle-aged men attempting a young man's  
sport. </MC></s>
```

```
<s><MC soa="ps"> He's not thinking about home runs anymore,  
just base hits. </MC></s>
```

```
<s><MC soa="ps"> Jaguar was shocked by Mr. Ridley's decision,  
because management had believed the government wouldn't lift the  
golden share without consulting the company first. </MC></s>
```

```
<s><MC soa="ps"> Some Democrats thought they might have  
compromised too much. </MC></s>
```

```
<s><MC soa="ps"> He savored the humor of his uncombed  
appearance among these buttoned-up chaps. </MC></s>
```

The tagger should remember to look carefully at the type of state being described; the clause type may be somewhat obscured by a colloquial wording, such as “right up their alley” or “headache”, which describe the attitudes held by the agents in these sentences:

```
<s><MC soa="ps"> Buying 51% of Rockefeller Group Inc. is  
right up Mitsubishi Estate Co.'s alley in one sense: The huge  
Japanese real estate company is entering a long-term relationship  
with a similarly conservative U.S. owner of tony urban property.
```

</MC></s>

<s><MC soa="ps"> That's a headache for long-term investors, since it forces them to reinvest their money -- usually at lower rates than the original mortgage securities carried. </MC></s>

It is also important, as with all sentence types, to look only at the contents of the main clause. The sentence shown here contains a clause that obviously describes a private state ("Desperately hoping"), but which is not in the main clause of the sentence, as the tag shows.

<s><MC soa="other"> Desperately hoping to spark sales, Nissan transferred 5,000 middle managers and plant workers to dealerships. </MC></s>

This subordinated phrase is a purpose clause, which will be discussed more in section 3.2.

Private states need not originate with human agents; other entities can be "experiencers" as well. These sentences are annotated with the comment "nh" to show that there is a non-human agent involved.

<s><MC soa="ps"> The nation's largest pension fund, which oversees \$80 billion for college employees, plans to offer two new investment options to its 1.2 million participants. </MC></s>

<s><MC soa="ps"> As the London trading session drew to a close, the market was still listening to the parliamentary debate on the economy, with new Chancellor of the Exchequer John Major expected to clarify his approach to the British economy and currency issues. </MC></s>

Several examples appear in the text of private states which originate with the reporter, rather than with an agent in the article, and because of this are presented to the reader in a more direct manner. Some are acknowledged by the reporter as his or her own private state, as in the following:

<s><MC soa="ps" fp"> We see an even stronger argument against UNESCO than its unsurprising failure to reform. </MC></s>

<s><MC soa="ps" fp"> I had to reach back to French 101 when the monsieur avec clipboard leaned over my shoulder during the coffee phase of dinner and asked whether I wanted to ride in a montgolfiere. </MC></s>

While a private state of the reporter need not always appear in first-person form, they will usually be presented in this manner. All these sentences have been annotated with "rps",

which stands for “reporter’s private state”, and with “fp” when in first person. This annotation should be used only when there is a private-state term with the reporter as the implicit experiencer. It should not be confused with the “rs” comment that recognizes reporter subjectivity, which can be used with many states of affairs, including those attributed to agents other than the reporter. For other discussion of how to tag first-person narrative, see section 4.4.

We will now make a bit of a digression into an interesting class of verbs that Quirk et al. call “seeming verbs.” The discussion will first show that they can be used either as private-state terms or as subjective elements, and will then show that many other words function like them in this respect.

The core members of this category are the following copular verbs: appear, feel, look, seem, smell, sound, taste. Depending upon what they modify, these verbs are either used to indicate how something is perceived by the senses, or something about the evidence one has for the information in a sentence. In these sentences, the seeming verbs are perceptual:

It felt cold and slimy.  
It looked beautiful.  
It smelled sweet.  
It sounded melodic.  
It tasted salty.

But in the following sentences, the seeming verbs say something about the source of information, mainly that the information is based on something that is observed. They also express some doubt.

She looked happy.  
(It is inferred from, say, her facial expression, that she is happy.)

It sounded like she had found the key. (It is inferred from the sounds that she earlier found the key.)

Similar words are the following, when they are used with inanimate subjects: indicate, make sense, mean, prove, reveal, suggest.

The clue indicated/meant/proved/revealed/suggested that the butler did it. (soa=“other”)  
The sudden flap of its tail proved it was still alive. (soa=“other”)

Now we get back to the purpose at hand. These words may or may **not** be private-state terms. It depends on how they are used in the sentence. If there is a “to clause” such as the ones in the following sentences, the sentences **are** about private states.

To John, Mary seemed happy. (soa=“ps”)  
The soup tasted great to John. (soa=“ps”)

The letter proved to him that Mary was the culprit. (soa="ps")

If there is no experiencer indicated, then the sentence is **not** about a private state. Instead, the verb is a "subjective element", or a word used to express belief, doubt, conjecture, or to otherwise convey the perspective of the reporter. Thus, these uses will be accounted for under the point-of-view tagging task. Note that one use of "feel" can be a private-state term even without a "to-clause". This meaning of "feel", shown in the second example that follows, is not a subjective element, because it acts more like "believe" than like "seem."

She felt terrible. (soa="ps")

She felt that Mary could have waited a little longer. (soa="ps")

Under the above uses of "felt", a "to John" clause is not possible (at least, it isn't possible without changing the meaning). When "feel" appears in the text, you will have to ask yourself carefully whether or not it is being used as a private-state term or as a "seeming verb".

The following verbs are similar to "seeming verbs", and there are others as well (more are shown below, but not all such verbs are mentioned in this document).

Hopefully, you will be able to generalize from the discussion here to other cases you find in the text. The verbs are: give the impression, make apparent, make clear, and show. The use of "show" that belongs here is "show" when used with an inanimate subject, for example, "It shows that crime does not pay."

If an experiencer is indicated, then these verbs are about private states. Otherwise, they are subjective elements (and should **not** be tagged as private states). In the following examples, they are subjective elements and not private states: jw-1-1-97: this part is wrong; look in trainingfile4.

Her bed showed the rumples of where her body had been. (soa="other").

He gave the impression of not exactly liking anything around the Hat Creek outfit.  
(soa="other"; the main clause is "He gave the impression")

His tone made clear that he was already tired of the subject. (soa="other"; the main clause is "His tone made clear")

If "give the impression", "make apparent", "make clear", or "show" are used with an **animate** indirect object, then they are not potential subjective elements, but instead are private-state terms:

The rumpled bed showed John that she had slept there last night. (soa="ps")

The clues made it apparent to John that the butler did it. (soa="ps")

As previously mentioned, subjective elements are in the realm of the point-of-view task, and the annotation "rs" signals a sentence with reporter subjectivity; this annotation is only for

reference with the point-of-view tagging task. jw12-16: I moved an example that was here because the tag isn't right, I think. Look for "hope seems doomed" to find it.

Here are two sentences from the Wall Street Journal containing words that work like seeming verbs with objects ("to Japanese investors", "to many"), and so are tagged as private states:

```
<s><MC soa="ps"> These securities are attractive to  
Japanese investors for three reasons. </MC></s>
```

```
<s><MC soa="ps"> To many, it was a ceremony more befitting a  
king than a rural judge seated in the isolated  
foothills of the southern Allegheny Mountains. </MC></s>
```

One use of "meant" also works like a seeming verb—it can either be a subjective element or a private-state term:

```
It meant that they were in trouble (soa="other")  
To John, it meant that they were in trouble (soa="ps").
```

Similar to these are sentences which attribute certain characteristics or states to agents in the text which are really, once again, the reporter's subjective statements, such as "does lend support" in the following:

```
<s> <MC soa="other"> The small increase in the index of  
leading indicators, which had climbed 0.5% in August but was  
unchanged in July, does lend support to the view that the economy  
has slowed noticeably. </MC></s>
```

jw12-16: (Although "view" appears, this sentence focuses on something providing evidence for something, which is "other" unless there is context to the contrary.) jw12-16: Julie – a purpose clause is something like "to go to the store", at least in Quirk et al's terminology. I changed the writing so we don't get confused.

Sentences whose main clauses are phrases like "was attempting to organize" or that contain "purpose clauses" such as "to help close the gap" in the examples below, are also to be tagged "other" rather than "ps". Such phrases focus more on the action of an agent in response to an internal state than on the state itself.

```
<s><MC soa="other"> Nomura was attempting to organize the 50  
billion-yen ($352 million) borrowing in Japan at a time when many  
Japanese banks, led by Industrial Bank of Japan, were pressuring the  
Wellington government to help them recover loans made to a defunct  
investment bank that had been owned by New Zealand's civil-service  
pension fund. </MC></s>
```

```
<s><MC soa="other"> To help close the gap, Nissan recently  
established a top-level cost-cutting committee. </MC></s>
```

The meaning of the words used should be carefully considered, however, to see whether they are focussing on the agent's response or the motivation for it. If the latter is the case, the sentence should be tagged as a private state. This example is a private state because the verb "seeks", though it is often used in purpose clauses to describe an action, has the meaning "wants" in this case.

```
<s><MC soa="ps"> The group also seeks uniformity among
states in provisions for taxpayers' rights. </MC></s>
```

Other clauses containing states of affairs that could be seen in some contexts as references to internal states, but which focus primarily on the action rather than the private state, should be tagged "other."

```
<s><MC soa="other"> Mr. Bush had been holding out for a bill
boosting the wage floor to $4.25 an hour by the end of 1991, coupled
with a six-month training wage for workers newly hired by any
employer. </MC></s>
```

jw12-16: Most actions are intentional and hence have a private state component, but we don't want to tag all actions as private states. The key is to ask what is the most focal component of the meaning. Holding out for a bill is similar to delaying or putting something off, which are more on the action/event side of things.

Following are four examples, for contrast, of sentences that jw12-16: describes things using the language of events and actions, but which focus on the private state that is manifested. This makes the private state the state of affairs to tag for. (The brackets added to the second case below, and elsewhere in the document, enclose a sentence that has been added not as a relevant example, but to provide the necessary context for an example.)

```
<s><MC soa="ps"> In addition, it's toying with an approach
that would impose health-expenditure ceilings or budgets on the
government as a whole and on individual states as a way to slow
health-care spending. </MC></s>
```

```
[<s> <MC soa="se.o"> Back in Berkeley, he was violently
scolded by a left-wing lady friend for consorting with such
people. </MC></s>]
```

```
<s> <MC soa="ps"> He became angry in return. </MC></s>
```

```
<s><MC soa="ps"> After 20 years of pushing labor proposals
to overhaul the nation's health-care system, Bert Seidman of the
AFL-CIO is finding interest from an unlikely quarter: big
business. </MC></s>
```

```
<s><MC1 soa="ps"> The court often weighs deductions of
```

sideline-business costs: </MC1><MC2 soa="other">  
Do they stem from a profit-seeking activity or a nondeductible  
hobby? </MC2></s>

Imperative sentences which contain a private state as their main clause's most significant element ("assume", "ask yourself", and "See"—a perceptual verb—in the examples below) should be tagged as private states (the imperative voice is often a form of reporter subjectivity, but does not negate the occurrence of the internal state that it describes, as will be discussed later):

<s><MC soa="ps"> Assume a neighborhood demonstration to  
protest speeding on a certain road or a careless accident involving  
a police car. </MC></s>

<s><MC soa="ps"> In either case, ask yourself whether you  
have become better informed on the issues under protest by watching  
the act of civil disobedience. </MC></s>

<s><MC soa="ps"> (See: "Revitalized Classics Take the  
Stage in Windy City" -- WSJ Oct. 19, 1989) </MC></s>

The focus and purpose of the reporter should always be considered, as a few words may make the difference in the type of the clause.

A final problem area, and hopefully an uncommon one, is illustrated by the following sentence, which at first glance appears to be a speech event of some kind.

<s><MC soa="ps" rps"> (I still say don't look down.  
</MC></s>  
<s><MC soa="other"> At least not when you are  
ascending.) </MC></s>

Although the use of the verb "say" normally implies a speech event, it is being used in this case to communicate the reporter's private state, which is equivalent to "I think (or believe) that you shouldn't look down." The use of first person speech by a reporter presents many difficulties such as this one, which do not typically appear in news writing. It is useful to note that the second clause in the above example, which is a sentence fragment, is tagged "other". This is the case with most fragments that appear in the text, and as shown here, a private state that is implied or projected from the preceding sentence ("I still say") is considered outside the realm of the main clause tagging task and is not applied to sentences that follow.

### 3.2.2 The Role of the Experiencer and other Topics

Note: As of dec 15, 1996, this section contains material that is up-to-date with respect to the tagging task, but contains a number of topics, and all of the topics may not fit under the title.

It is necessary to discuss in more detail the relationship between private states and their experiencers, as this is often a difficult relationship to see, and as it has a great deal to do with classifying sentences as private states. In addition, separate from the issues of experiencers, the structure and the wording of some sentences shows clearly that an internal process or motivation has occurred or is occurring, but in other sentences they almost seem to hide it. The understanding of a clause as a private state frequently depends on its main verb, and this section will discuss in particular several verbs that present challenges to the process of tagging private states. (As you will see, sometimes you will have to consider other parts of the sentence than the main verb, because the main verb itself is too “weak” to stand on its own.)

Some verbs that accompany private states (and speech events, at times) have an influence on the content of the sentence that is easily “weakened” by the appearance of states of affairs which are more likely to command the reader’s attention. The verb “make”, along with verbs having similar meanings, is such a verb. Although it is hard to imagine “make” itself describing an internal state of an agent, a sentence such as “John made Mark think” brings a definite state of affairs to mind: Mark is thinking. While John is not physically forcing Mark to undergo a mental process, he acts as the motivation for that process, which is what the sentence focuses on. Here is a Wall Street Journal sentence with this use of “make” and a private state:

```
<s><MC1 soa="ps"> Please make me and thousands of other
readers more comfortable with our daily purchase of your
newspaper by raising your environmental standards to your
overall impeccable quality levels, </MC1><MC2 soa="other">
and increase your use of recycled paper. </MC2></s>
```

Before considering some more examples, let us give some advice to the tagger. The issues being discussed point out that the tagger must sometimes consider more of the sentence than the main verb, and, in fact, might tag the sentence based on other parts of the sentence. When does this occur? When the main verb, as it is being used in that sentence and context, is too “weak” to stand on its own. We appeal to the tagger’s intuition here. Start by considering the main verb. Do you know enough about what kind of event or state it is to stop and tag based on that? If not, go on to the next highest-level thing in the sentence. Note that you are trying to get at what kind of event or state it is. Once you are able to determine this, stop – do not go on to consider the arguments of that action or state. It must be stressed that this issue is being discussed under the topic of private state, in this version of the instructions. But this advice and this general issue is relevant for the other classes too.

Below are some examples of sentences that have a verb like “make” as their main verb, and that present private states as their primary focus.

```
<s><MC soa="ps"> The pocketbook impact of health
benefits has driven business and labor to a surprising
consensus. </MC></s>
```

In the “driving a car” meaning of “driving”, you would only need to consider the main verb. You would have a good idea what kind of event is being described—that use of “driving” has enough substance to stand on its own. But “driving” here does not stand on its own – it is like “causing” or “making”. We need to look at more of the sentence. “business and labor” don’t tell us what type of event or state is being described. To see what type of state or event it is, you need to consider: “Drive ⟨some agents⟩ to a consensus”. Now you see it is a private state. Another strategy is to paraphrase the sentence: “the impact made them come to the same opinion” is a rough paraphrase that reveals the private state. (Note that “consensus” could be used in a speech event sentence, in another context and with a word like “come”: “They sat down and talked and came to a consensus.” But the context of the above sentence does not highlight a communicative action of coming to a consensus, but rather focuses on the fact that they have come to the same opinion.)

```
<s><MC1 soa="other"> The politician appearing is highly
controversial </MC1><MC2 soa="ps"> and has recently
generated a good deal of rancor amid certain groups. </MC2></s>
```

“Generated” in this sentence does not stand alone. What was generated? Discussion could have been generated, say, in which case the tag would be se.o; or even energy could have been generated, in which case the tag would be “other”. Here, “rancor” is generated, which is a private state.

Here is another description of the above examples: verbs such as “generate” and “driven” in the above examples direct the focus of the sentence from the originator of the action to its recipient, and to the resulting private state.

Here are two similar examples for which intuition is key:  
jw12-16: note to julie and karen – this one is from the diffs file, number \*29\*. Karen – I couldn’t find the example you referred to in the diffs file that we said was other but you said was ps. Please feel free to ask if/when you find it.

```
<s snum=1 s_id=891101-0141_7_1> <MC soa="ps"
rs"> Such a countermove could end Jaguar’s hopes for remaining
independent and British-owned. </MC></s>
```

“Ending” isn’t enough to consider in itself, so we go further in the sentence. “Ending hope” is enough; we see it is a private state. Here is a similar one:  
jw12-16: I moved this from above, where it had the tag other.

```
<s><MC1 soa="ps" change in data"> Mr. Mayor’s hope
that references to "press freedom" would survive
unamended seems doomed to failure; </MC1>
<MC2 soa="se.ds"> the current phrasing
is "educating the public and media to avoid
manipulation." </MC2></s>
```

Consider the first conjunct. In this sentence, “seemed” is a subjective element, so we disregard it for this tagging task. The future and the negativity of “doomed” are also disregarded: they concern the subjectivity of the reporter’s predictions. Thus, when we get to the core of the state of affairs type, this one is very similar to the one above: it is equivalent for our purposes to “His hopes failed”, which is like the “hope ending” example above.

jw12-16: The most obvious cases in which you have to look further than the main clause are probably those with “to be” or “seemed” as the main clause. Clearly, these do not stand on their own. So, “She was happy” is a private state, as is “She seemed happy” (because we disregard the subjective element “seemed”, these two sentences are equivalent for our purposes).

There is a great deal of syntactic variation among the types of sentences for which you have to look further than the main verb. We direct the tagger to rely on his or her intuitions, and we do not attempt to lay out a set of syntactic rules to follow. jw12-16: But, for the syntactically inclined, it may be helpful to note that “She was happy” has a Subject-Verb-Complement structure, and the “driven” sentence above has a Subject-Verb-Object-Complement structure. In the first case, the complement is an adjective. In the second case, the complement is expressed by a prepositional phrase (“to a consensus”) instead of by a single word, as it would be in a sentence like “He drove her crazy”. Recognizing that “to a consensus” is the attitude adopted by the object (“business and labor”) makes it clear that this sentence depicts a private state.

The verbs “compelled to” and “forced to” also belong in this category of possibly “weak” verbs, as used, for example, in the following sentence:

```
<s><MC soa="se.o"> Currently, the rules force executives,  
directors and other corporate insiders to report purchases and  
sales of their companies' shares within about a month after the  
transaction. </MC></s>
```

Notice that “force” could be used with a much stronger meaning, in which case it would stand on its own – if the context sets things up so that someone were being physically forced through a door, for example, we would tag for the main verb, and the tag would be “other”. Here, you are tagging for “force ⟨someone⟩ to report ⟨something⟩” — hence, the se.o tag (for “report”).

In contrast, here is a sentence with a similar structure to the ones being considered here, but without a private-state meaning. The phrase “threw additional burdens on...” might seem comparable to the verbs above, and in a sentence like “The events burdened their consciences” it would in fact have a private-state meaning. As it appears below, it does not focus on an internal state but rather an external situation involving actions that must be performed.

```
<s><MC1 soa="other"> Their cowardly acts of civil  
disobedience, which they tried to hide under the cloak of outrage  
at a war they characterized as "immoral," weakened the national  
fabric </MC1><MC2 soa="other"> and threw  
additional burdens on those who served honorably in that conflict.
```

</MC2></s>

As always, it is important to consider the meaning of the sentence and the words it contains, in the particular context in which they appear.

Here is a more difficult case, which also uses the verb “generated”:

```
<s><MC1 soa="other"> The ads have generated some laughs
</MC1><MC2 soa="ps"> but also plenty of attention
because they are so unlike any other U.S. auto advertising.
</MC2></s>
```

While the first clause of this sentence describes an action (laughing), the state that is the focus of the second clause, “attention”, is an attitude that would be classified as a private state. The fact that the sentence does not specify an experiencer does not prevent this classification; the SVOC structure mentioned above is not the only syntactic possibility for private state sentences of this type. Private states with nonspecified experiencers will be discussed at a later point. It is important to note, incidentally, that the word “attention” puts the internal state it describes at the center of attention because it is the head noun in its clause.

The importance of context to the classification of a sentence is seen once again in the preceding example. Such a sentence can be tagged as a private state even though it presents no agent as the experiencer of that state, because the existence of the agent can be determined from two factors: the fact that the private state does exist, and the appearance or implication of potential agents in the surrounding text. Several verbs will be discussed next that commonly help to describe private states in sentences where their experiencers may not be made clear.

The previous discussion of “seeming states” made the distinction, among sentences that point to private states, between those that indicate an experiencer for the state (which are likely to be tagged “ps”) and those that do not (which are likely to be tagged “other”). The presence of a “to clause” specifically naming the experiencer is an obvious contributor to a decision in favor of a private state tag, but such helpful hints are not always provided by the text. Therefore, for private-state terms whose experiencers **could** be indicated in a sentence, (by naming the subject, using a “to clause”, for example) but are used **without** an explicit experiencer, the following rules will hold:

- A. If the default experiencer is the reporter, the private-state term itself does not make the tag ‘ps’

The sentence might be tagged ‘ps’ for another reason, but NOT because of this term.

‘It seemed pretty’ (tag is other)  
‘He seemed angry’ (tag is ps, due to ‘angry’, not ‘seemed’)

- B. If the default experiencer is not the reporter, the sentence should

be tagged as a private state. Such sentences may take one of the following forms: ‘‘It was intended to...’’, ‘‘It is thought that...’’, ‘‘It is beloved’’, ‘‘It is meant to...’’, ‘‘It was designed to...’’, ‘‘It is believed to be...’’ (and others as well).

The term ‘‘default experiencer’’ simply means the entity who, in the absence of any special context indicating otherwise, is recognized as the experiencer of the state mentioned. The reporter is the default experiencer in the case of most seeming states, because they are attitudes that originate with and are expressed by him or her rather than by other agents in an article. The reason we specify ‘‘by default’’ rather than making a black and white statement is that sometimes there is an implicit experiencer of the sentence. For example, ‘‘John looked at the stone. It seemed pretty.’’ The second sentence presents John’s reaction, set up by the first. Recognizing such sentences belongs to the subjectivity tagging task, not to this one. Because the default experiencer of ‘‘seemed’’ in ‘‘it seemed pretty’’ is the reporter, and ‘‘being pretty’’ is tagged ‘‘other’’, the sentence should be tagged ‘‘other’’.

Here are two sentences using the word ‘‘pressure’’ to describe private states. The first supplies an explicit experiencer (‘‘dessert-menu planners’’), and the second requires a look at how the sentence is being used, which shows that the default experiencer is not the reporter. (The context for the second example is supplied in brackets, and the sentence in question appears at the end.)

```
<s> <MC1 soa="other"> Still, the restaurant’s ever-changing menu
of five-course dinners -- <MC2 soa="other"> it supposedly
hasn’t repeated a meal since opening in 1971 <\MC2> --
requires constant improvisation. <\MC1> </s>
<s><MC soa="ps"> And that puts added pressure on
Chez Panisse dessert-menu planners. </MC></s>
```

```
[<s><MC soa="other"> A bipartisan commission established
jw-1-1-97: why isn’t this se.o -- presenting new plans?
by Congress and headed by Sen. John Rockefeller (D., W.Va.)
is scheduled to present new plans for dealing with the uninsured
and long-term care for the elderly by next March 1. </MC></s>
<s><MC soa="ps"> A quadrennial commission
appointed by Health and Human Services Secretary Louis Sullivan
is taking a broad look at the economics of Medicare for the elderly,
Medicaid for the poor and the health system in general. </MC></s>
<s><MC soa="ps"> It is expected to report next summer.
```

```
</MC></s>
```

```
</p>
```

```
<p>
```

```
<s><MC soa="se.ds"> "No magic bullet will be discovered next
```

year, an election year," says Rep. Fortney Stark (D., Calif.)  
 </MC></s>  
 <s><MC soa="other"> But 1991 could be a window  
 for action. </MC></s>]  
 <s><MC soa="ps"> The pressure for change will rise  
 with costs. </MC></s>

It should be noted that “pressure” is not by nature only a private-state term; in contexts in which speech is predominant, it can be used to describe a speech event rather than a private state (and there may even be uses that can be tagged as “other”).

Another group of verbs to which the rules given above must be applied includes some uses of “meant”, and some uses of perceptual verbs like “smelled”. As stated earlier in the discussion of private states, these verbs are often used as subjective elements, so an experiencer must be present in order to call them private-state terms. Below is a comparison of two states expressed with and without experiencers.

"John smelled the soap" (private state)  
 "The soap smelled good" (other, with subjectivity).

"John meant to use the spoon to eat." (private state)  
 "The spoon is meant to be used to eat" (other; not subjective).

The following sentences do not include experiencers. The default experiencer is **not** the reporter, so they **should** be tagged as private states.

<s><MC1 soa="other"> The golden share dates from Jaguar’s  
 public offering in 1984 </MC1><MC2 soa="ps">  
 and was designed to protect the company from takeover.  
 </MC2></s>

<s><MC soa="ps"> The index is intended to measure future  
 economic performance. </MC></s>

jw12-16: (I moved these examples here, and changed the tags from ps to other – the reporter is by default the experiencer) And, in this example, the experiencer is not included and **is** by default the reporter, so the tag in this case is “other”:

<s><MC soa="other" change in data">  
 The case of Rosa Parks, the black woman  
 who refused to sit at the back of the bus, comes to mind as an  
 illustration. </MC></s>

<s><MC soa="other" change in data">  
 It felt more like a party, or a  
 highly polished jam session with a few friends, than a classical

concert. </MC></s>

Here is some tagging advice for the case when you have what might be a private state term in a sentence and the experiencer is NOT specified. To decide if you have a subjective element or a private state, take the sentence out of its context, and add the reporter as the experiencer. (This is the one place in these instructions where we will tell you to take a sentence out of context to make a decision.)

If the meaning changes greatly, the private-state term is **not** being used as a subjective element. Wouldn't you be surprised to find out that the reporter is the one who happened to design the share to protect the company, or that it is the reporter who intends the index to measure performance, if nothing said so? The reporter saying that something was designed in a certain way does not suggest that the reporter designed it. In the second case, the index is intended to measure future economic performance presumably by those who created it.

Here are some made-up examples where a word **is** a subjective element:

“The recent increase in robberies means that our social fabric is decaying.” (subjective element; tag is “other”)

(It means this to the reporter, without any special context.)

Remember, the test is whether the reporter is the experiencer by default. Consider:

“In J. Smith's book, he takes a dim view. The recent increase in robberies means that our social fabric is decaying.”

You may read the second sentence as expressing J. Smith's view, that the robbery rate means **to Smith** that our social fabric is decaying. But remember that we are leaving such questions to the subjectivity tagging task. Your job is to: take the sentence out of context, and ask if the reporter is by default the experiencer. If ‘yes’, then the element is a subjective element, and you should not tag for it. It will be up to the subjectivity taggers to note that the subjective element belongs to Mr. Smith, not to the reporter.

### 3.2.3 Nonspecific Experiencers

With some private-state sentences, it is left unclear who is experiencing the state. This type of sentence deserves special attention, because it is not always obvious that these cases are private states like the others, and it is also difficult at times to determine who the experiencer is. Often, the experiencer is the reporter, who seems to try to add credibility to his or her reaction to an event by attributing it to a more general and implied audience of experiencers.

The term “nonspecific” is used when the reporter refers to an implied, impersonal experiencer or group of experiencers of a private state or, occasionally, originators of a speech event. At times the reporter may not know enough about these agents to specify who they are; in other cases he or she seems to be trying to add credibility to statements, involving these agents, which are actually expressions of his or her own opinions. The point of view

in these sentences is usually a subjective one because of this. In the following sentence, the reporter uses the pronoun “one”, a common choice for referring to a nonspecific experiencer, to indirectly express an opinion about the value of a book: jw12-16: I moved the following sentence from just above to down here near the other “one” sentence.

```
<s><MC soa="ps"> Nonetheless, one can't help  
wonder whether the game will be just a little bit slower.  
</MC></s>
```

```
<s><MC1 soa="ps"> One learns a lot from this book,  
<MC2 soa="ps" nse"> or seems to, </MC2> about crippling  
federal bureaucracy. </MC1></s>
```

To interpret MC2, you have to reconstruct the sentence, filling in the components that are missing due to ellipsis. Thus, MC2 is equivalent to “One seems to learn a lot from this book”. Now, we see that “seems” is a subjective element, so MC2 is the same, for our purposes, as “One learns a lot from this book”. This receives the “ps” tag for “one learns”.

The next example shows another common pronoun, “you”, which is used in the same way as “one” in sentences of this type.

```
<s><MC soa="ps" rs"> You'd think all the stories about  
well-heeled communities and developers getting HUD grants would  
prompt Congress to tighten up on upscale housing  
subsidies. </MC></s>
```

jw12-16: Note that we are treating these experiencers as explicit experiencers, even though they are rather non-explicit. So, the sentence “It seems to one that...” would be tagged a private state, because there is an explicit experiencer (albeit a vague one).

### 3.2.4 Partial List of Private-State Terms

We close this section with a partial list of private state terms, some of which were also discussed above. This list is not intended to be exhaustive. In fact, such a list is probably not possible.

jw12-16: The tagger is free to skip this section if they desire.

#### Psychological Terms

##### Psychological Adjectives

afraid, alert, angry, ashamed, aware, certain, conscious, curious, dying  
(e.g., “He was dying to tell Lippy about it”),

eager, embarrassed, delighted, exasperated, frightened, frustrated, glad, happy, hurt, indifferent, indignant, impatient, jealous, loath to, miserable, proud, outraged, relieved, reminded,

sad, satisfied, scared, shattered, sorry, sure, surprised, taken aback, tired, undecided, unhappy, used to

(e.g., “He was used to the noise”),

wary, worried.

Two kinds of private-state sentences with psychological adjectives are subject-complement (SC) sentences:

John became angry.(soa=“ps”)

John was unhappy.(soa=“ps”)

and subject-verb-object-complement (SVOC) sentences:

He made John angry.(soa=“ps”)

The experiencer is the subject in the SC sentences and the object in the SVOC sentences. Subject or object complements with “mood” also denote private states:

He was in a bad mood. (soa=“ps”)

It put him in the mood to celebrate.(soa=“ps”)

Two of these adjectives, “certain” and “curious”, are subjective elements rather than private-state terms, if the thing modified is a noun phrase about something inanimate, or is a clause. In the following two sentences, they are private-state terms:

The sound made John curious. (soa=“ps”)

Mary was certain that John had been there. (soa=“ps”)

But in the next two sentences, they are subjective elements:

It was certain that it was about to rain. (soa=“other”)

The fact that she didn’t send the letter was curious. (soa=“other”)

## Psychological Verbs

**(A)** The experiencer is subject in a subject-verb-object (SVO) sentence (e.g., “John decided to leave”), or the object in an subject-verb-object-complement (SVOC) sentence if the psychological verb appears in the complement (e.g., “The letter made him decide to leave”).

believe, care

(e.g., “He cared about Mary”),

check oneself

(e.g., “He was about to inform her but he checked himself”, but not, e.g., “He checked himself

out of the hotel.”),

consider, decide, despise, dwell  
(e.g., “He had begun to dwell on the incident”),

expect, figure  
(e.g., “He figured that she wanted to go”),

forget, feel  
(e.g., “He felt sad” or “He felt that John was asking too much”)

find  
(find oneself, e.g., “He found himself in a large room”; the special use of “to find” in, e.g., “He awoke to find Mary gone” but not in “He went to the library to find Mary”, where “to find” is a purpose clause; “find out”; or, e.g., “He found it incredible”; but not, e.g., “He found a sock”),

hate, hope, imagine  
(e.g., “He imagined himself sitting in a Porsche” or “He imagined that she would be there”),

know, like, love, mean  
(i.e., intend, e.g., “He meant to find out”),

notice, realize, regret, remember, say to oneself, see  
(e.g., “He saw that she was unhappy”),

sigh to oneself, suppose, tell that  
(e.g. “She could tell that he was angry”),

think  
(“think that”, “think of”, “think about”, and just “think”, without an object),

tire, understand, want, wish, wonder, would rather, would sooner, yearn.

**(B)** The experiencer is the object in an SVO sentence (e.g., “It surprised him”), or the indirect object in an SVOO sentence (e.g., “It surprised him that Mary was here”; “him” is the indirect object and “that Mary was here” is the direct object). Note that the past participle forms of most of these verbs correspond to psychological adjectives listed above, e.g., “He was surprised” (soa=“ps”). Also, the **present** participle forms of some of these verbs correspond to adjectives that are subjective elements, e.g., “It was surprising.” (soa=“other”). Others correspond to subjective adjectives that have other forms, e.g., “bothersome” and “scary” which correspond to “being bothered” and “being scared”, respectively. (“It was bother-

some/scary” is “other”; “It was bothersome/scary to John” is “ps”.)

bother, cheer, come  
(as in, e.g., “A silly old hymn came to him”),

exasperate, frighten, frustrate, get to  
(as in, e.g., “The noise got to John”)

remind, satisfy, scare, shame, stop  
(when the subject is not something that **physically** stops the object; e.g., “Gus’s remark stopped him”, but not “The barrier stopped him from entering the room”),

strike  
(e.g., “It struck him that she was in love with him”),

surprise, touch, worry.  
(as in e.g., “She touched him with her kindness” but not as in, e.g., “She physically touched him.”)

## **Psychological Nouns**

astonishment, composure, delight, feeling  
(e.g., “The feeling stayed with him all evening”, “He had a bad feeling”, “He rode with a good feeling in his heart”),

happiness, hatred, hope, misery, realization, sense, spirits  
(e.g., “His spirits fell when he read the letter”),

thought, urge, yearning.

## **Perceptual Terms**

### **Perceptual Verbs**

hear, see  
(Note a special use of “to see” to denote a private state: “He awoke to see nothing familiar”; but not, e.g., “He went to the store to see a movie”, where “see” is in a purpose clause),

smell  
(e.g., “He smelled the coffee brewing”).

### **Perceptual Nouns**

glimpse, vision

## Seeming-State Sentences

Recall from above that seeming verbs are subjective elements. Examples are “seem”, “appear”, and the perceptual verbs “look”, “feel”, “smell”, “sound”, and “taste” when the way that something being perceived looks, feels or sounds is described. Others are “mean”, “prove” and “show” when used with inanimate subjects. For example:

He sounded/looked/seemed happy. (soa=“ps” due to “happy”)

It looked/felt/appeared as if it was going to rain. (soa=“other”)

She appeared/seemed to be happy. (soa=“ps” due to “happy”)

She looked pretty. (soa=“other”)

She touched him. (soa=“other”) He felt cold. (soa=“other”)

The telltale sign meant/proved that John had been there. (soa=“other”)

However, as discussed above, if an experiencer is indicated, as in the following:

It looked to Call as if it was over. (soa=“ps”)

It felt to Newt like velvet. (soa=“ps”)

He sounded/seemed/looked/appeared happy to Mary.(soa=“ps”)

It proved/meant to him that she did not love him any more.(soa=“ps”)

then the state of affairs denoted by the sentence is a private state.

As with other private-state sentences, seeming verbs can be followed by private states that have non-human experiencers:

```
<s><MC soa="ps" rs"> It appears that the only thing  
Congress is learning from the HUD story is how to enlarge its  
control of the honey pot going to special interests. </MC></s>
```

## Other Private-State Terms

Here are some private-state sentences that do not fit into the above categories.

She shivered in spite of herself (soa=“ps”)

(this is a use of “shivers” that focuses on the experience of shivering)

He had a plan in mind

She was dizzy with fatigue

She smiled/sighed to herself

Of course, since metaphor can be used for private states, there is a wide variability in private-state sentences. Most, if not all, private-state sentences that employ metaphor contain some psychological or perceptual term. Here are two examples:

The girl was clutching her composure to her heart.  
Zoe's heart bounced.

### 3.3 “Other” Clauses

The largest category of those being recognized in this task is the one given the tag “other”. Most often, these are simply statements, of fact or of opinion, made by the reporter. They do not introduce any agent into the text, and they do not attempt to present any point of view or record any internal state or occurrence of speech. This classification includes, among other things, listings of statistics or other data, and actions performed by an agent that are not primarily communicative or indicative of a private state. Some examples of sentences whose main clauses are tagged “other” are the following:

```
<s><MC soa="other"> In asset terms, Mitsubishi Estate is  
the largest real estate firm in Japan. </MC></s>
```

```
<s><MC soa="other"> BELL INDUSTRIES Inc. increased its  
quarterly to 10 cents from seven cents a share. </MC></s>
```

```
<s><MC soa="other"> In the outfield, Paul Blair, the Orioles'  
eight-time Gold Glove winner, elegantly shags a fly. </MC></s>
```

```
<s><MC soa="other" ls"> PRIME RATE: 10 1/2%.  
</MC></s>
```

```
<s><MC1 soa="other"> Today, the pixie-like clarinetist has mostly  
dropped the missionary work (though a touch of the old Tashi still  
survives) </MC1><MC2 soa="other"> and now goes on the road with  
piano, bass, a slide show, and a repertoire that ranges from light  
classical to light jazz to light pop, with a few notable  
exceptions. </MC2></s>
```

Here are two examples of “other” clauses containing statements of opinion from the reporter (note the comment tag “rs”):

<s><MC1 soa="other"> It was enjoyable to hear  
accomplished jazz without having to sit in a smoke-filled club,  
</MC1><MC2 soa="other"> but like the first half, much  
of it was easy to take and ultimately forgettable. </MC2></s>

<s><MC soa="other"> The Babelists of the United  
Nations are experts at obfuscation. </MC></s>

Note that while “hear” and “take” in the first example are perceptual verbs, which in other contexts would be tagged as private states, jw12-16: in this context they are tagged “other” for reasons already given in previous sections. First, “enjoyable”, “easy”, and “forgettable” are subjective elements, so we aren’t considering them for this tagging task. There is no explicit experiencer in the sentence, and the default experiencer of “hear” and “take” in these sentences is the reporter. In addition,

Thus, we are left with the tag “other”.

Other examples of this clause type (“other”) may consist of comments injected into a discussion by the reporter:

<s><MC soa="other"> What’s next? </MC></s>

<s><MC soa="other" sf"> No way. </MC></s>

The verbs used in many “other” clauses are similar or identical to verbs used elsewhere to describe speech events. These may include “appointed”, “named”, “quoted”, “vetoed”, “submitted”, “indicted”, “charged”, “ruled”, “introduced (a bill)”, “resigned”, and “denied”. The examples below illustrate some common uses of such verbs in “other” clauses:

<s><MC soa="other"> Robert Q. Marston, president emeritus,  
University of Florida, and a director of this maker of medical  
devices, was named chairman. </MC></s>

<s><MC soa="other"> A three-judge panel of the court in Cincinnati  
made the ruling Saturday. </MC></s>

<s><MC soa="other"> DRUG WARS: A Texas legislator proposes  
color-coding drivers’ licenses of some drug offenders. </MC></s>

<s><MC soa="other"> State Senator J.E. "Buster" Brown, a Republican  
who is running for Texas attorney general, introduced the  
bill. </MC></s>

<s><MC1 soa="other"> He claimed losses totaling \$42,455 --  
</MC1><MC2 soa="other"> and the IRS denied them all. </MC2></s>

<s><MC soa="other"> And a Fannie Mae seminar  
this week promises to draw hundreds of prospective investors,  
who can be expected to channel tens of billions of dollars into  
the market in the next few years. </MC></s>

jw12-16: In all of these cases, the focus is more on an action than on a speech event, even if some of them involve writing.

Here are some “other” sentences that might resemble private states:

<s><MC soa="other"> He has been considered several times for  
appointments to federal district and appellate court vacancies in  
Pennsylvania. </MC></s>

<s><MC soa="other"> Japan has found another safe outlet  
for its money: U.S. home mortgages. </MC></s>

<s><MC soa="other"> But there is a risk, which the Japanese  
discovered when they first dipped their toes into the market  
nearly five years ago. </MC></s>

Again, the focus is not on a private state to a large enough extent to warrant a “ps” tag.

It is also worth noting again that playing music and singing, while they may be seen to have a communicative component, are included in the “other” category since their primary purpose is not to present a point of view:

<s><MC soa="other"> Then, as if to show that he could play fast  
as well, he offered the second movement from Saint-Saens’s Sonata  
for Clarinet, a whimsical, puckish tidbit that reflected the flip  
side of the Stoltzman personality. </MC></s>

<s><MC soa="other"> Glamorous and pure-voiced as ever,  
Ms. Collins sang Joni Mitchell’s “For Free” -- about an encounter  
with a street-corner clarinetist, to which Mr. Stoltzman contributed  
a clarinet obligatto -- and Mr. Douglas’s lush setting of a Gaelic  
blessing, “Deep Peace.” </MC></s>

## 4 Other Problem Issues

Some distinctive issues arise in this task which are not related directly to any of the main clause categories, but whose effect on the tagging deserves some notice. Some, such as reporter subjectivity (“rs”) and implicit private states (which are also known as “represented thoughts”), will be dealt with later in the point-of-view tagging task, and are included in the main clause tags only for reference with that task. Others, such as nested compound

sentences, represent areas related to this tagging task which may later be examined in more detail. The following sections examine some of these issues.

#### 4.1 Indicating Unusual Cases in the Tags

Several annotations can be added to the main clause tags to point out occurrences of the issues mentioned above. These are possible values of the “” field.

##### Main Clause Tag Annotations

```
aut citation of the article's author
cs nested compound sentence appears within a main clause
fp first-person speech
hd headline
nh non-human agent
ls non-sentence listing of data
mca segmentation of main clauses is ambiguous
nse non-specific experiencer
rps reporter's private state (as opposed to an agent's)
rs reporter subjectivity
rt represented thought (implicitly a private state of an agent in the text)
sc semicolon that separates subordinate rather than main clauses
sf sentence fragment
ws written speech event
wsj comments made directly by the Wall Street Journal
```

Implicit private states, or “represented thoughts”, are recognized in this task only by the above annotation, and are not a major concern in tagging main clauses, except that recognizing an implicit state (which will usually be tagged “other”) at times saves the tagger further debate over its type. For reference, then, here is an example of represented thought:

```
<s><MC soa="ps"> Institutional investors that must evaluate
poison pills on a regular basis are interested less in this general
debate than in the answers to specific questions about the
corporation issuing the pill. </MC></s>
<s><MC soa="other"> Does this corporation have a
high-quality management team with a good track record?
</MC></s>
<s><MC1 soa="other"> Does this team have a viable
strategy for improving shareholder values, </MC1>
<MC2 soa="other"> and does this strategy require
implementation over an extended period? </MC2></s>
```

The second, third and fourth main clauses in this example are given the comment “rt” because they represent the “questions” of investors, mentioned in the first sentence. One issue that stands apart from the others in this section is the treatment of various comments made by the Wall Street Journal, which follow some articles. Such comments are noteworthy because they are a rare exception to the fact that the reporter’s speech occurs at the highest level of the text. The following is an example of this kind of sentence, which appears at the end of some articles and supplies information about the reporter who wrote it:

```
<s><MC soa="other"> Mr. Shupe is co-author (with David  
G. Bromley) of "'Moonies' in America: Cult, Church, and Crusade" and  
"Strange Gods: The Great American Cult Scare." </MC></s>
```

Such sentences will usually be tagged “other”, and are annotated with “wsj” to show that they did not originate with the reporter as did the rest of the article preceding them.

## 4.2 Nested Compound Sentences

This task does not attempt to assign separate tags to the individual clauses of a compound sentence if the sentence, in its entirety, is understood to be the object of an explicit speech event or private state. Although such nested compound sentences may, in a later tagging task, be broken up and tagged separately, they are not treated as independent sentences in the current task. However, the tagger should note these cases by including the annotation “cs” in the tag. This annotation should not be used for normal compound sentences, but only for those that are understood as the object of another main clause.

```
<s><MC soa="se.ds"> "I think it would be irresponsible  
and I am not about to be blackmailed by Japanese banks or any other  
international interests." </MC></s>
```

The objects of the speech event in the sentence is described by a compound sentence: the first conjunct is the span from “investors” to “proposals”, and the second is the span from “and” to “memory”. Such a compound sentence is subordinate to the main clause – “[agent] said [speech]” – even if this clause does not occur explicitly in the sentence, as in the second example above.

### 4.2.1 Focus of Attention and Context Dependence

Just as this task focuses more on the main clauses of sentences than on the information that occurs outside their boundaries, it also recognizes that the arrangement of the text according to the reporter’s style and purposes results in some states and events being brought more into focus than others. The following examples might have been tagged as speech events in other contexts, because of their use of the verbs “quoted” and “promises”, but because of the role of these verbs in the text have been placed in the “other” category.

<s> <MC soa="other"> A seat on the Chicago Board of Trade was sold for \$350,000, down \$16,000 from the previous sale last Friday. </MC></s>

<s> <MC soa="other"> Seats currently are quoted at \$331,000, bid, and \$350,000, asked. </MC></s>

<s> <MC soa="other"> And a Fannie Mae seminar this week promises to draw hundreds of prospective investors, who can be expected to channel tens of billions of dollars into the market in the next few years. </MC></s>

For contrast, here are four “se.o” sentences whose main verbs, “lobbying”, “agreed”, “resurfaced”, and “paint a picture” might in some cases be seen as actions, but in these examples were judged to place sufficient focus on speech events:

<s> <MC soa="se.o"> Now UNESCO apologists are lobbying President Bush to renege on President Reagan’s decision to depart. </MC></s>

<s> <MC soa="se.o"> Armstrong World Industries Inc. agreed in principle to sell its carpet operations to Shaw Industries Inc. </MC></s>

<s><MC soa="se.o"> Differences between the Treasury and the Federal Reserve on the usefulness of intervention to help restrain the dollar resurfaced at the hearing. </MC></s>

[<s><MC soa="other"> Last March, nine months after the judge’s swearing-in, the state attorney general’s office indicted him on a sweeping array of charges alleging more than 10 years of "official oppression" in Cambria County, a depressed steel and mining community in western Pennsylvania. </MC></s>]

<s><MC soa="se.o"> The allegations, ranging from theft and bribery to coercion and lewdness, paint a disquieting picture. </MC></s>

Sentences should be tagged as “se.o” instead of “other” only if a speech event is such a sufficiently large part of the meaning and the focus of the sentence that it “crosses the line” into the realm of speech events, and is seen as the state of affairs that the sentence describes. Most, if not all, actions have an intentional component that may seem to conflict with the main action being described—all speech, for example, involves private states to some degree. The goal of this task, however, is to target the type of the “largest” or most intentional meaning component of the main clause. This next sentence was tagged “se.o” because even though it describes an action, it was seen to emphasize the speech component of the action,

and also because classifying it as speech is consistent with sentences such as “He had no comment”, which will be discussed later.

```
<s><MC soa="se.o"> Mr. Wakeman, whom Chemical tried to
keep, didn't return calls for comment. </MC></s>
```

Many times the decision of what is focal in a sentence depends on the context in which that sentence appears. The second sentence in the example below shows a private state that is context-dependent.

```
<s><MC soa="se.o"> Back in Berkeley, he was violently scolded
by a left-wing lady friend for consorting with such people.
</MC></s>
<s><MC soa="ps"> He became angry in return. </MC></s>
<s><MC soa="ps"> He had developed a hatred for the hacker and a
grudging appreciation of the federal "spooks" who make national
security their business. </MC></s>
```

If followed by a speech event sentence, the phrase “became angry” might describe speech—that is, something that was said in anger. In this case, however, the sentence following shows that “angry” describes the agent’s emotional reaction, which is a private state. See the two accompanying training files for more such examples.

Following are three sentences whose verbs often describe speech, but that refer to actions in the context in which they appear.

```
<s><MC soa="other"> FIRST AMERICAN FINANCIAL Corp.
declared a special dividend of one share of Class B common stock for
each share of Class A common stock, payable to holders of record on
Nov. 10 if the Securities and Exchange Commission approves this as
the effective date of the registration statement. </MC></s>
```

```
<s><MC soa="other"> Under the guise of "healing the wounds of the
nation," President Carter pardoned thousands of draft evaders, thus giving
dignity to their allegations of the war's "immorality." </MC></s>
```

```
<s><MC soa="other"> Mr. Bush had been holding out for a bill
boosting the wage floor to $4.25 an hour by the end of 1991,
coupled with a six-month training wage for workers newly hired
by any employer. </MC></s>
```

jw12-16: Shall we just delete the following paragraph? We decided (in the “diffs” file) that both of the “defined” in what follows are “other”. The decisions on the following two sentences were influenced by the context provided by the article they appeared in and by the tenses of the verb phrases in both sentences. The first sentence was tagged “se.o”, because, like the surrounding discourse, it relates a situation, a series of diplomatic talks, in which speech

predominates. With the shift in tense in the second sentence, however, “defined” becomes more stative, describing the way things are instead of the way they are talked about.

```
<s><MC soa="se.o should be other"> Five of the six categories of weapons
under negotiation haven't even been defined. </MC></s>
<s><MC soa="other"> Tanks currently are defined as
armored vehicles weighing 25 tons or more that carry large
guns. </MC></s>
```

The following example illustrates another context-related decision. While the verb “dictate” would in many contexts imply a speech act, a closer look at the type of event being described here shows that no actual speech is involved, but rather that the reporter is simply making a statement about the nature of an understood scientific concept:

```
<s><MC soa="other"> The laws of heredity dictate that half of the
plants springing from these greenhouse-produced seeds will be male
sterile and herbicide resistant and half will be male fertile and
herbicide susceptible. </MC></s>
```

As shown, the sentence should be tagged “other”. The next example contains descriptions of both a speech event (“arguments”) and a private state (“led the president to...”). This is an illustrative example of the influence that the reporter’s focus can have in a certain context.

```
<s><MC soa="ps"> Though President Bush's political
critics at home have been urging him to open a more direct
dialogue with Mr. Gorbachev, it actually was the arguments of
leaders within the Soviet bloc itself that led the president to
seek the December meeting. </MC></s>
```

Although technically the main clause of this sentence is “it was the arguments”, a speech event, the reporter has used this type of syntax just for style. The most important part of the sentence is the president’s internal state, described in the word “led”. This is even more evident when the context, which refers to several internal states of the president, is examined. The sentence is tagged “ps”. Many times the decision of what is focal in a sentence is even more dependent on the context in which that sentence appears. The second sentence in the example below shows a private state that is context-dependent.

### 4.3 Sentence Fragments

The appearance of sentence fragments in a number of articles has caused some inconsistency in the tags, since each line of text was previously labelled with sentence markers, and is tagged in this task as if it contains at least one independent clause. The function of sentence fragments differs with the writer of the article; at times they have a practical purpose, as in these examples which are part of listings of data within an article (these are annotated with “ls”):

<s><MC soa="other" ls"> PRIME RATE: 10 1/2%.  
</MC></s>

<s><MC soa="other" ls"> @ Index of lagging indicators  
..... 119.7 -.5 </MC></s>

<s><MC soa="other" sc, ls"> CERTIFICATES  
OF DEPOSIT: 8.09% one month; 8.04% two months; 8.03%  
three months; 7.96% six months; 7.92% one year. </MC></s>

Other instances reflect a certain flair on the part of the reporter, or a particular style or mood that he or she is trying to create in the article. jw12-16: I hate to point this out, because it will hurt the system, but aren't the second two below speech events too? Isn't there ellipsis that has to be carried on (I'm talking landing in a canal in a porous wicker basket, etc) I am confused about our tagging of sentence fragments – do we try to “fill out” the sentences, or not?

Conclusion from later discussion (12-17): ellipsis and anaphora should be resolved in tagging the sentences. Thus, all of the following three tags are se.o.

<s><MC soa="se.o"> I'm talking about landing in a  
canal. </MC></s>  
<s><MC soa="se.o"> In a porous wicker basket.  
</MC></s>  
<s><MC soa="se.o"> With a pilot who speaks no  
English. </MC></s>

Still other examples of sentence fragments can be found in the text surrounding an article, such as the headline (annotate with “hd”), or in citations of authors (annotate with “aut”). Most sentence fragments will simply be tagged “other”, as very few have been found which involve speech events or mental states at this point. Fragments that are found to describe speech events or private states, as the second part of the following example does, should be tagged as such.

<s><MC soa="other" hd"> 'Frequent Drinker' Offer </MC></s>  
</p>  
<p>  
<s><MC soa="se.o" hd"> Stirs Up Spirited Debate </MC></s>

(In a novel one finds private-state sentence fragments, such as “The awful pain”.) The “sf” annotation is important because it recognizes text that might otherwise be excluded from the tagging because it does not contain a proper main clause. On completion of this task, every line of text should be tagged as if it were a sentence, and the annotation “sf” added whenever necessary to show that a line of text is a fragment instead of an actual sentence

as the sentence markers imply. An additional point to remember on the subject of sentence fragments is that, like other sentences in the main clause task, they are being tagged for their event type and not for implicit states or points of view that may be “projected” onto them from preceding or following sentences. Any such outside influences will be recognized in the point of view tagging task.

#### 4.4 First Person Speech

First person narrative, while it occurs infrequently in the text, shows the potential for some difficulty. The direct presentation by a writer of a point of view, such as a private state or a subjective statement, is a departure from the normal writing style found in the text, and is not consistent with the expectations of this task about the way that information is given to the reader. In the text being tagged, first person speech comes in two forms. A reporter may write an article in first person in order to set a less formal tone or to describe an event in which he or she was personally involved. The first example below is from such an article. A reporter, or a reader of the Wall Street Journal, may write an editorial article or a letter to the paper’s editors. The second sentence below is from an editorial article.

```
<s><MC soa="ps" fp"> I had to reach back to French 101
when the monsieur avec clipboard leaned over my shoulder during the
coffee phase of dinner and asked whether I wanted to ride in a
montgolfiere. </MC></s>
```

```
<s><MC soa="ps" fp"> We see an even stronger argument
against UNESCO than its unsurprising failure to reform. </MC></s>
```

Sentences containing occurrences of first-person speech are marked with the annotation “fp”, and should receive main clause tags that correspond as well as possible to the clause type represented, although these may not be as clearly distinguished as they are in other sentences. In general, the tagger should keep in mind that this is a unique category of sentences which may require slightly different treatment by the tagger.

#### 4.5 Sentence Types Related to Subjectivity

The task described in this document largely focuses on a sentence’s semantic meaning, such as one would find in a dictionary, rather than what a particular sentence **actually** describes or denotes. Of course, it is difficult to tease these two issues apart entirely, as we have seen the influence of context on deciding whether, for example, “agree” is being used to describe speech or a private state. But we are asking which meaning the terms have (in context), rather than about properties of the particular state or event being referred to.

This leads us to the perhaps counter-intuitive categories of speech events and private states discussed next: speculated, negated, and suggested states and events. These three sentence types, which are closely related, involve different types of conjecture about a particular speech event or private state, or the negation of it, by the reporter. All these deviations

in the presentation of states and events result from the use of some degree of subjectivity by the reporter writing the sentence. Since subjectivity is often a cause of difficulty in the recognition of main clauses, it will be helpful to spend some time looking at these ways in which subjectivity is manifested in the various sentence types. More information about point of view can be found in the instructions for the point-of-view tagging task.

#### 4.5.1 Speculated States and Events

Subjectivity, as previously defined, involves the expression of evaluation, judgment, belief, or conjecture. As evaluation or judgment, it might take the form of a sentence like the following (in the phrase “downright comic”):

```
<s><MC soa="other"> Some town-watching excursions  
were downright comic. </MC></s>
```

A sentence containing conjecture, or speculation, is shown in this example, in the phrase “is likely to be interrogated”:

```
<s><MC soa="se.o"> BRAZILIAN INTERROGATION  
Brazilian financier Naji Nahas, who was arrested on Monday after  
102 days in hiding, is likely to be interrogated next week by the  
Brazilian judiciary. </MC></s>
```

Here is a more complex example of speculation:

1. 

```
<s><MC soa="se.o"> The first thing anybody will tell you  
about ballooning is that it requires zip in the way of athletic  
prowess, or even a measure of derring-do. </MC></s>
```
2. 

```
<s><MC soa="se.o"> They will also tell you that even if  
you hate heights, you can still balloon. </MC></s>
```
3. 

```
<s><MC soa="ps"> (I still say don't look down. </MC></s>
```
4. 

```
<s><MC soa="other"> At least not when you are ascending.)  
</MC></s>
```
5. 

```
<s><MC soa="se.o"> What they won't tell you is not to go  
aloft in anything you don't want to get wet. </MC></s>
```

The first, second, and fifth clauses in this paragraph are predictions, rather than actual reports, of speech. A point-of-view analysis of these sentences would indicate that the speech mentioned hasn't actually happened, and that it is therefore the reporter's subjective statement that it will occur. However, when the focus is kept on the type of event of the main clause, the sentences must be seen as being about speech events, since that is the type of event they describe. Observations on whether this speech event has taken place yet, or will ever take place, are within the realm of the point-of-view tagging task. For the main clause task, the sentences specified should be tagged “se.o”, as the main clauses are talking about speech events that involve paraphrase, and they can be annotated with “rs”. Here is an example of a private state that is given this type of treatment:

<s><MC soa="ps"> Do they want the spotlight for themselves or for their cause? </MC></s>

Although the private state described by “want” is included in an unresolved question asked by the reporter, its existence is assumed by the reporter, as the question being asked is actually “They want the spotlight, but do they want it for themselves or for their cause?” The conjecture-related subjectivity in the sentence is related not to the goals of the party in question, but their reasons for those goals. Therefore the sentence, whose main clause *does* concern the goals, should be tagged “ps”. More examples of this type of sentence are included below.

<s><MC soa="ps"> If the answers to these questions are affirmative, then institutional investors are likely to be favorably disposed toward a specific poison pill. </MC></s>

<s><MC soa="ps"> But despite the acrimony between Mr. Ross and Mr. Yetnikoff, officials of the Time side of Time Warner have reportedly been increasingly interested in a settlement that might yield attractive business opportunities. </MC></s>

The verbs describing private states in the preceding sentences are “be favorably disposed” and “been increasingly interested”.

#### 4.5.2 Negated States and Events

States or events that are presented as not happening are to be treated, in the main clause tagging, in the same way as states or events that do occur. This task’s goal of capturing the main idea that the sentence is trying to present requires that a private state or a speech event that is described, whether in an affirmative or a negative sense, should be tagged as that state or event. Negation is simply a modifier, just as the adverb “not” in a sentence such as “She is not happy” modifies the adjective “happy”, which is a description of the internal state of the sentence’s subject. Whether the sentence affirms or denies the happiness of the subject, it is still about her emotional well-being. Here are some examples of negated private states:

<s><MC soa="ps"> When the jolt is timed just right, the subjects don’t recall seeing the first group of letters. </MC></s>

<s><MC soa="ps"> A company spokesman didn’t know Mr. Wakeman’s age. </MC></s>

<s><MC soa="ps"> Not all those who wrote oppose the changes. </MC></s>

Negated speech events, while often harder to classify, are treated in much the same way. The term “lack-of-speech event”, while not an official category in the tagging, has been used as an attempt to categorize a smaller group of sentences within the “se.o” classification in

which speech events are referred to specifically because they do *not* happen. Such sentences are included in the category described in section 3.1.3 because they refer to, describe, or otherwise introduce the idea of a speech event's occurrence, but usually do not present its contents. These sentences further group themselves together, however, because they negate the occurrence of the speech event, or at least of some part of its content that might have been expected to occur. While sentences in this category are less likely to have a subjective focus than those containing speculation, the presentation of a state or event in a negated form might still be prompted at times by some subjective purpose of the reporter. Examples of this sentence type would be the following:

```
<s><MC soa="se.o"> Terms weren't disclosed. </MC></s>
```

```
<s><MC1 soa="other"> Mr. Agnelli and the Aga Khan also have some  
business ties, </MC1><MC2 soa="se.o"> and a spokesman for the  
Agnelli company didn't rule out that the current agreement could  
lead to further collaboration. </MC2></s>
```

```
<s><MC soa="se.o"> I'm not referring to the traditional  
champagne drenching during the back-on-terra-firma toast. </MC></s>
```

```
<s><MC soa="se.o"> The White House is purposely not  
calling the meeting a summit so that there won't be any expectation  
of detailed negotiations or agreements. </MC></s>
```

```
<s><MC1 soa="se.o"> But word of the possible session was closely  
held by the president and a handful of top aides, </MC1><MC2  
soa="se.o"> and word of it didn't reach many second-level officials  
until the past few days. </MC2></s>
```

```
<s><MC soa="se.o"> In the wake of British Chancellor of the  
Exchequer Nigel Lawson's surprise resignation and sterling's  
subsequent nose-dive, most analysts had little good to say about  
the pound's near-term prospects. </MC></s>
```

Note that in the fourth sentence in the preceding group, the word “purposely” is a signal of a purpose clause (see 3.2) rather than a private state, so that even though the sentence speaks of an agent's motives or intentions, it should be tagged for its main state of affairs, which is a speech event or “se.o”.

#### 4.5.3 Suggested States and Events

Sentences closely related to negated and speculated types are those in which the reporter proposes a state or event that could possibly occur, but for which there is no agent or actual occurrence in the text.

```
<s> <MC soa="ps"> Lawyers and their clients who frequently  
bring business to a country courthouse can expect to appear before  
the same judge year after year. </MC></s>
```

Sentences with this type of subjective focus often use the auxiliary verb “can” to describe the agent’s optional role in the sentence. In the main clause task, the preceding sentences are treated as private states, just as they would be if written “He dismissed him as a lightweight” or “lawyers expect to appear...” in accordance with the goal of tagging for the state of events described in the sentence. The next example shows an even closer resemblance to sentences containing speculation:

```
<s> <MC soa="se.o"> The precedent having been set,  
who can complain if future generations called upon to defend the  
U.S. yield to the temptation to avoid the danger of combat by simply  
declaring the war immoral and hiding until it is over? </MC></s>
```

The possibility of the speech event described (“can complain”) depends on the event described in the “if” clause that follows it (“yield to the temptation”). This sentence, like other speculated and suggested forms, is to be tagged as the speech event or private state which it describes—in this case, as “se.o”.

Karen: This isn’t that important but I think it’s a new section to the last version of instructions

#### 4.6 Mistakes in the Original Text or its Formatting

A problem that faces the tagger occasionally, and at times may impede the tagging process, is the occurrence of mistakes in the original text that is being tagged. When found, and when necessary, such errors can be corrected, as long as any changes to the original copy of the text are carefully recorded, and are duplicated in any copies of the text that have been made. The main reason for the correction of errors is that in a few instances they have prevented the proper tagging of the sentence for its understood state of affairs. An example is the following, which appeared in the original copy of the Wall Street Journal text being tagged:

```
<s> This is no idle fearlast year the FHA lost $4.2 billion in  
loan defaults. </s>
```

When corrected and tagged, the sentence read as follows:

```
<s><MC1 soa="other"> This is no idle fear: </MC1>  
<MC2 soa="other"> last year the FHA lost $4.2 billion in loan  
defaults. </MC2></s>
```

The sentence could actually have been corrected in several different ways, since the method used to separate the two clauses cannot be determined by looking at the original, but a colon was judged to be one of the most likely ways that the sentence would have been divided. While

this is not a problem area that should consume a great deal of time or effort, proper treatment is very important when the problem does occur. When an error concerns the division of the text into sentences, however, the error cannot be corrected, but only adapted to. Below is a case in which the formatting of the text conflicted with the goals of the main clause tagging:

```
<p>
<s><MC soa="other" hd"> Trial and Terror
</MC></s>
</p>
<p>
<s><MC soa="ps" sf"> At times I sequester my mind
</MC></s>
</p>
<p>
<s><MC soa="ps" sf"> When I must think with precision,
</MC></s>
</p>
<p>
<s><MC soa="ps"> Detached from all other thoughts
</MC></s>
</p>
<p>
<s><MC soa="other"> While trying to reach a decision.
jw1-1-97: this should be ps.
</MC></s>
</p>
<p>
<s><MC soa="other"> But often nothing's resolved,
</MC></s>
</p>
<p>
<s><MC soa="ps" sf"> To my frustration and fury:
</MC></s>
</p>
<p>
<s><MC soa="other"> With pros and cons in limbo,
</MC></s>
</p>
<p>
<s><MC soa="ps" sf"> I feel like a hung jury.
</MC></s>
</p>
<p>
```

```
<s><MC soa="other"> -- Arnold J. Zarett.
</MC></s>
</p>
```

This example is a short poem whose lines have been tagged as though they were individual sentences, even though the poem itself is composed of a single long sentence with several clauses. This formatting causes further complications by dividing the “sentences” of the poem into paragraphs, which eliminates the option of tagging it as a compound sentence. The solution in this case is to tag each line as the main clause type it most closely corresponds to, and to add the “sf” comment to show that each clause is really a fragment. Note that a result of this is that several of the clauses above, when viewed individually, can be tagged as private states. In this, as in other cases, all sentence and main clause segmentation presented in the text should **not** be altered, but tagged as it appears.

\*\*\* jm1021: There’s this one, too, but I didn’t know what to say about it. Should it be included in this section on mistakes in the text?

```
<s><MC soa="se.o"> BRAZILIAN INTERROGATION
Brazilian financier Naji Nahas, who was arrested on Monday after
102 days in hiding, is likely to be interrogated next week by the
Brazilian judiciary. </MC></s>
```

\*\*\*

#### 4.7 Borderline and Ambiguous Cases

A very small group of sentences was found that is not easily included under any one of the categories defined in this task. The difficulty in tagging these sentences comes from the fact that a single main verb is used to describe two different states of affairs simultaneously, and to say that one is significantly more focal than the other would be to disregard part of the meaning of the sentence. Most often, this problem is found in sentences that simultaneously describe speech events and private states. Within the goals of this task, it is difficult, if not impossible, to determine which of these states, events, or attitudes such verbs describe. If a single type of state of affairs cannot be determined, such sentences can be given a tag that accounts for both states of affairs at once. The same scheme will be used in all of the tagging tasks:

| Tag | Definition   |
|-----|--|
| --- | -----  |
| QZ  | unknown tag for an object.                                 |
| X Y | borderline between X and Y. X, Y are tags which are known. |
| X?Y | ambiguous between X and Y. X, Y are tags which are known.  |

jm12-14: Are we still including the "qz" tag in this task? jw12-16: Let's get rid of the qz tag.

jw12-16: And let us say that, in freezing these instruction and data at this time, there should only be borderline cases of ps—event, and only in the cases described below. A later tagging task might allow more ambiguity.

For example, if it cannot be determined from the context whether a sentence describes a speech event or private state, it should be given the tag "<MC soa="se.o?ps">". jm12-14: Would this tag, which is only hypothetical at this point, be written "ps?event" to be consistent with the current borderline tag? jw12-16: Let's address this question after the current task; it is best in this study that there aren't any of these – we should make the tagger choose, except in the case described below.

If it is on the borderline between the categories "se.o" and "ps", or between "ps" and "other"—that is, if the state of affairs described can be seen as both a private state and another event at the same time, and neither more clearly than the other—the tag should read

<MC soa="ps|event">.

To restate, then, a sentence is tagged as ambiguous when repeated readings do not resolve the question of which category it belongs to, and as a borderline case when, given its context, it cannot be pinned down as a member of any single category to which it belongs more than all the others. It is very important, in both cases, to consider the sentence's context carefully before choosing either a borderline or an ambiguous tag, as the proper category for most sentences can be determined by looking at their surroundings if the sentences themselves are not easily classified. A truly ambiguous sentence is usually the result of poor writing on the part of the reporter, in which sufficient context is not provided in which to understand the sentence's intended purpose. A truly borderline sentence occurs only when the choice of words (usually the verb) and the surrounding text create a main clause that clearly serves a dual purpose, according to the system of classification used in this task, and the two categories that it represents are equal in their importance.

jw12-16: The current borderline cases require yes'es to all three questions below, and the tag should be ps|event: (I just want the same criteria I stated in the boderline file):

1. Is the type of state of affairs being tagged for an event or events (rather than a state)?
2. Do the event or events involve interaction between agents, or some other strong private state? AND
3. Are you not sure that all of the actions are speech events?

Here is another statement, in terms of three criteria that must be true?

1. we are tagging for an event (or events)
2. we know they are interactional events
3. but we cannot say that the actions are all speech events (or else we know that they aren't)

So these are cases of events that strongly involve a ps, and we cannot say that they are all speech events. Calling them ps wouldn't work, because they are events; calling them speech event wouldn't work, because other kinds of actions are involved; and calling them "other" wouldn't work, because then we would lose the interaction/private state completely. So, we use the new tag. Please see the training files accompanying these instructions for more examples and description of these cases.

The tendency of some verbs to express borderline states of affairs varies greatly with their context as well as their tense; potentially borderline sentences should be given a very careful reading, and the type and the purpose of all surrounding sentences considered. Here are all some sentences that were tagged as borderline in this task:

```
<s><MC soa="ps|event"> Mr. Blair and Hees have been feuding
for months. </MC></s>
```

```
<s><MC soa="ps|event"> Eastern's creditors committee,
along with the company, has consistently opposed the
pilots' claim, which if paid would have to come out of
money both hope to use to pay off other bankruptcy
claims. </MC></s>
```

```
<s><MC soa="se.o|ps"> But they have been at odds over how
much Mr. Hunt would owe the government after his assets are sold.
</MC></s>
```

```
<s><MC1 soa="se.o"> Eastern has disputed the claim,
</MC1><MC2 soa="ps|event"> but a federal district court,
an appeals court and now the arbitrator have all sided with the
pilots. </MC2></s>
```

```
<s><MC1 soa="other"> Not only are Japan's financial institutions
putting their enormous clout to work; </MC1><MC2 soa="ps|event"
> increasingly they're squaring off against one another in
unprecedented public fashion. </MC></s>
```

```
<s><MC soa="ps|event"> The NAM embraces efforts, which
both the administration and the medical profession have begun, to
measure the effectiveness of medical treatments and then to draft
medical-practice guidelines. </MC></s>
```

```
<s><MC soa="ps|event"> The AFL-CIO also embraces treatment
guidelines. </MC></s>
```

Verbs that have meanings similar to the preceding ones, but that are less obvious signals of either speech events or private states, may also be used to describe both states of affairs

simultaneously:

```
<s><MC1 soa="other"> Both sides are jealously guarding their
turf, </MC1><MC2 soa="ps|event"> and relations have been at a
flashpoint for months. </MC2></s>
```

```
<s><MC1 soa="ps|event" fp, rs"> I am not a proponent of
everything Gandhi did, </MC1><MC2 soa="other"> but some of
his law breaking was justified because India was then under
occupation by a foreign power, and Indians were not able to
participate fully in decisions that vitally affected them. </MC2></s>
```

The second main clause in the first example is similar to sentences with “at odds”, which imply the involvement of both communication and attitudes, and is tagged “ps—event” to show that it represents both. The second example describes a state-being a “proponent”—that implies actions such as speaking out for someone and voting for them, and also attitudes such as believing in and supporting. Neither can be understood independently of the other. Note the use of the verb “triggered” in the next two examples. This is a “weak” verb, of the kind described earlier, which puts the borderline states of affairs in these sentences (“clash” and “opposition”) into focus:

```
<s><MC soa="ps|event"> What triggered the latest clash was
a skirmish over the timing of a New Zealand government bond
issue. </MC></s>
```

```
<s><MC soa="ps|event"> Any major attempt to revamp the
health-care system is likely to trigger opposition from politically
powerful interest groups, particularly the American Medical
Association, and perhaps from the public as well, if Congress takes
steps that patients fear will limit the availability of
care. </MC></s>
```

Below are some cases that could be confused with borderline sentences, as they can be seen to have private state and speech elements. In each case, the state of affairs represented in the tag was seen to be the one that was more focal. jw12-16: I changed the se.ds tag in the following, in accordance with the borderline file.

```
<s><MC soa="se.ms" ws"> "If HUD is to be reformed,"
it concluded, Members of Congress will "have to start looking into,
and doing something about, the practices of their colleagues."
</MC></s>
```

```
<s><MC soa="se.ds"> "The oldest designer got to work on the
dashboard," she recalls. </MC></s>
```

<s><MC soa="se.ms"> He expects such a cut "because of problems in several sectors of the economy, particularly real estate and automobiles." </MC></s>

Note that these sentences were identified and classified in the same way as the previous examples, with their main verbs being the primary determining factor.

Sentences with “agree” are particularly varied in the interpretation they can receive. These examples were tagged as either speech or private states because of their respective contexts.

<s><MC soa="se.o"> People close to the GM-Jaguar talks agreed that Ford now may be able to shut out General Motors. </MC></s>

<s><MC soa="ps"> The two sides don't even agree about how much money is at issue. </MC></s>

<s><MC soa="se.o"> Mr. Achenbaum, who had been considering paring down his firm or merging it with another small consulting outfit, soon agreed. </MC></s>

<s><MC soa="se.o"> Ryukichi Imai, Japan's ambassador to Mexico, agrees that Mexico may be too eager. </MC></s>

<s><MC1 soa="ps"> Agreement on these points is a long way from a specific program, </MC1><MC2 soa="ps"> and nobody expects the U.S. to rush toward radical restructuring of the health-care system. </MC2></s>

In contexts using the sense of “agree” that involves the making of a deal, treaty, or business transaction, this verb should be tagged as a speech event. Such a context puts the focus on the actions and words of the participants rather than on their internal states. A sentence containing “agree” which is surrounded by speech events or puts the focus on speech is also more likely to belong with speech-only sentences than with those in the borderline category. The “ps” classification represents an understanding that the attitude described by “agree” is more important than the speech event that manifests it. Following are some sentences with this particular “make a deal” use of “agree”.

<s><MC soa="se.o"> Regarding Mr. Hunt's taxes, he and the IRS have apparently agreed on a basic formula for liquidating his estate in which the IRS would get 70% of the proceeds from a liquidating trust and 30% would go to other creditors. </MC></s>

<s><MC1 soa="se.o"> The IRS had demanded \$90 million

</MC1><MC2 soa="se.o"> but Mr. Hunt would agree to no more than \$60 million. </MC2></s>

<s><MC soa="se.o"> The Aga Khan, meanwhile, agreed to trade some of his stake in Luxembourg-based Ifint S.A., another Agnelli family company, for 7.45% of Giovanni Agnelli & Co.'s capital. </MC></s>

<s><MC soa="se.o"> Instead, Sony is likely to agree to let Warner participate in certain of its businesses, such as the record club of Sony's CBS Records unit. </MC></s>

<s><MC soa="se.o"> In a forward-rate agreement, a client agrees to an exchange rate on a future currency transaction. </MC></s>

<s> <MC soa="se.o"> Congressional Democrats and the Bush administration agreed on a compromise minimum-wage bill, opening the way for the first wage-floor boost in more than nine years. </MC></s>

<s><MC soa="se.ms"> In their new contract this year, American Telephone & Telegraph Co. and the Communications Workers of America agreed to look for "prompt and lasting national solutions" to rising health-care costs. </MC></s>

Karen: This is a new comment, "mca", for ambiguous segmentation, not classification, of sentences. I don't know if it's come up in the problems files you've worked on.

Even fewer ambiguous sentences have been found in the tagged text, and the problem they present cannot easily be resolved using the formula mentioned at the beginning of this section. These sentences, whose segmentation into main clauses is ambiguous, are almost impossible to tag in a way that represents all of the multiple interpretations that are possible. Partly for this reason, sentences in this task are separated into their component main clauses as part of the preprocessing, rather than asking the tagger to make what would be difficult judgment calls. If sentences of this type are found (and they will be rare), the comment "mca" can be added to the clause or clauses to show that there is ambiguity about how the sentence should be read and segmented. As with other decisions in sentence classification, the tagger should choose the tags for the type and comment fields that are the most likely choices, and that can be the most reasonably defended according to the goals of this task. Following are some examples of sentences that posed this problem of ambiguity.

<s><MC soa="se.o" mca"> The dispute over that issue, according to one U.S. official, is a "potential treaty stopper," and only President Bush and Soviet leader Mikhail Gorbachev may be able to resolve it. </MC></s>

The ambiguity in this example comes from the fact that its second clause could either be seen as part of the speech event, or an “other” statement outside of it. The interpretation chosen was the former, and the tags represent the view that the entire sentence is a speech event with more than one clause. The following are two more examples in which the ambiguity concerns the number of tags that might be given as well as their type. It would be possible to read each of these sentences either as a speech event with a compound sentence as its object, as they are presently tagged, or as an “other” statement originating with the reporter (“investors...had until today” and “the administration doesn’t want...”) and an se.o statement of another agent (“the issue has produced...Mr. Lane said” and “He said it hasn’t yet been determined...”). If either were viewed as the latter, it could be tagged as a regular compound sentence, which would require the addition of another main clause tag. The choice made here was to call both sentences speech events, as their tags show.

```
<s><MC soa="se.o" mca"> Investors, money managers and
corporate officials had until today to comment on the proposals, and
the issue has produced more mail than almost any other issue in
memory, Mr. Lane said. </MC></s>
```

```
<s><MC soa="se.ms" mca"> He said it hasn't yet been
determined how the RTC will raise the cash, but the administration
doesn't want it to be included on the federal budget, because it
would "distort" the budget process by requiring either exemptions from
Gramm-Rudman or big increases in the budget deficit. </MC></s>
```

Here is a similar sentence for which the other alternative, recognizing two clauses instead of one, was chosen:

```
<s><MC1 soa="other"> Nissan handled the die-hards in a typically
Japanese fashion: </MC1><MC2 soa="se.ms"> They weren't fired but
instead "were neglected," says Kouji Hori, the personnel manager at
the Nissan Technical Center. </MC2></s>
```

Here is a slightly more difficult example.

```
<s><MC1 soa="se.o" mca"> Nissan concedes that it won't
recoup all its market-share losses in Japan until at least 1995,
</MC1><MC2 soa="other"> and even that timetable might
prove optimistic. </MC2></s>
```

This example presents a problem similar to the preceding ones: the second clause could be seen as either the reporter’s subjective statement, or as part of the statement made by Nissan, in which case it would be understood as a represented thought. In this sentence, then, either of two very different interpretations would be expressed and distinguished only by what could be placed in the comment fields of the tags. If the example is understood as the first situation described, the comment would only read “nh” in the first tag, and “rs” in the second; if

seen as the second case, the comment would be “nh, cs” (recognizing the nested compound sentence) in the first clause’s tag, and “rt” in the second (meaning that the second clause is interpreted as being implicitly the speech event of Nissan). Although the comment tags are not the most important part of this task, this example shows how subtle differences in the interpretations of a sentence are sometimes represented in its comment tags.