# Simple Metadata Annotation Specification

# Version 5.0 – May 14, 2003

**Stephanie Strassel**

**(strassel@ldc.upenn.edu)**

**Linguistic Data Consortium**

# 1 Introduction

Working with pre-existing verbatim transcripts of broadcast news and conversational telephone speech, annotators will identify several types of *metadata*. The ultimate goal of metadata annotation is to support the production of transcripts that are maximally readable. This document defines the Simple Metadata task. SimpleMDE defines a subset of the full range of metadata phenomena. Under SimpleMDE, annotators will identify fillers, depods (the *de*letable *po*rtion of an edit *d*isfluency), and SUs ("semantic" units). Transcripts annotated for metadata can be "cleaned up" for readability; for instance, depods and fillers might be removed and each SU presented as a separate line within the transcript. Each metadata type is described in detail in the sections below.

Two types of data will be annotated: broadcast news and conversational telephone speech. The data types are different from one another in a number of ways. Broadcast news data is single channel data, meaning that all speakers' voices are presented in a single audio signal. Broadcast news comprises read speech, formal interview format, man-on-the-street interviews and some spontaneous speech, though usually not conversational. Telephone speech is two channel data, so that each speaker in the phone call is presented on a separate audio signal. Telephone speech is spontaneous, and much of it quite informal and conversational. Broadcast news will usually be simpler for metadata annotation. Annotators can expect to see fewer fillers and disfluencies. With respect to SUs, annotators can expect to see quite a difference in the prosody and discourse structure of broadcast news data, particularly read speech, when compared to conversational telephone speech. Intonation and breath patterns are likely to be very different in read speech and sentences will be structured differently. This may affect the placement of some SUs.

For both data types, annotators will perform metadata annotation using both the audio file and a verbatim transcript, executing multiple passes over the data and using a tool customized for each specific annotation task. For telephone speech annotators will annotate one speaker at a time, but will always have access to both audio channels as well as the complete 2-channel transcript. Acoustic signal information is encoded during initial transcription of the audio file. The first metadata annotation pass focuses on filler and depod annotation, and the second pass identifies SUs. Additional passes over the data that focus on quality control are conducted to improve annotation accuracy and consistency.

## 1.1 A word about notation

Within this document, the following notation conventions are used.

***Fillers and Edit Disfluencies***

| | |
|---|---|
| [ text ] * | ‘text’ is a depod and ‘*’ is its interrupt point. |
| <u>text</u> | ‘text’ is the corrected portion of the disfluency |
| **text** | ‘text’ is a filler (DM, FP, Aside/Parenthetical or EET) |

| | |
|---|---|
| te- | ‘te’ is a partial word token |

***SUs***

*Sentence-level SU Breaks*

| Symbol | Description | Shorthand |
|---|---|---|
| **/.** | Statement SU break | Statement SU |
| **/?** | Question SU break | Question SU |
| **/@** | Backchannel SU break | Backchannel SU |
| **/-** | Incomplete SU break | Incomplete SU |

*Sentence-internal SU Breaks*

| Symbol | Description | Shorthand |
|---|---|---|
| **/&** | Coordination SU break | Coordination SU |
| **/,** | Clausal SU break | Clausal SU |
| ∅ | No break (where one might be expected) | |

# 2 Fillers

We will consider four kinds of fillers: filled pauses (FP), discourse markers (DM), explicit editing terms (EET) and asides/parentheticals (A/P). None of these filler types alters the propositional content of the material into which they are inserted, and their insertion does not depend on the word identities of the surrounding material. Annotating fillers consists of identifying the token(s) that act as a filler, and labeling the filler with the appropriate type.[1] Cleanup of these phenomena is equivalent to removing the entire word or phrase that constitutes the filler. Each filler type will be considered in turn below.

## 2.1 Filled pauses

Filled pauses are hesitation sounds that speakers employ to indicate uncertainty or to maintain control of a conversation while thinking of what to say next. Filled pauses do not add any new information to the conversation (other than to indicate the speaker's hesitance) and they do not alter the meaning of what is uttered. For instance,

**Um** I do **uh** some **uh** woodworking myself.

Filled pauses can occur anywhere in the stream of speech. Each language has a limited set of filled pauses that speakers can employ. For English, this set includes the following five items:

### English Filled Pauses

[1] Note that each filler also contains an interruption point at the left edge of the filler's extent. Annotators are not required to explicitly label this filler IP; this information will be added automatically by the annotation tool. Further note that when the interruption point from an edit and filler coincide, they will share the same (single) IP tag, as in this example:

[red] * **uh** blue

Since they are not manually annotated, these left-edge filler IPs are **not** reflected within this document.

ah uh
eh um
er

Other sounds or non-lexemes can occasionally be used as a filled pause, and some speakers may adopt an idiosyncratic filled pause noise that does not appear on the above list. For the purposes of SimpleMDE annotation, we limit ourselves to the filled pauses listed above.

**Note**: The annotation tool pre-identifies filled pauses (limited to *ah, eh, er, uh, um*) and automatically pre-annotates them as filled pauses. Pre-identified filled pauses are displayed in blue font; annotated filled pauses are displayed with blue underlining. Annotators must verify that each pre-labeled filled pause candidate is actually acting as a filled pause, and must remove the annotation from any non-filled pauses.

Be aware that some tokens that can be used as FPs may have other functions, like question responses and backchannel cues, elsewhere in the discourse. Label tokens as filled pauses only when they indicate a speaker's hesitation.

### 2.2 Discourse markers

A discourse marker (DM) is a word or phrase that functions primarily as a structuring unit of spoken language. DMs frequently appear at the beginning or end of an SU. To the listener, a DM signals the speaker's intention to mark a boundary in discourse, such as a change in the speaker, the beginning of a new topic or the expression of a response:

```
Yeah, that gets on my nerves too.  So anyway, tell me about your
new job.
```

DMs can also serve to indicate the speaker's attitude or orientation toward the discourse; for instance, a speaker may introduce a discourse marker to indicate a contradictory stance toward what the other speaker has stated:

```
A:  I think he's done a terrible job in that position.  If it
were up to me I'd fire him.
B:  See, I don't know if I'd go that far.
```

Because of the many uses of DMs in speech, and the resulting complexity of defining and identifying them, we will annotate only a limited set of discourse markers that are used in clearly recognizable ways. The following words and phrases will be annotated as discourse markers *when they are used to structure the discourse and do not carry separate meaning*.

| | |
|---|---|
| **actually** | **now** |
| **anyway** | **see** |
| **basically** | **so** |

| | |
|---|---|
| **I mean** | **well** |
| **let's see (now)** | **you know** |
| **like** | **you see** |

Some words not contained on this list may function as discourse markers, and the list may be expanded in future task definitions, but for this version of SimpleMDE discourse markers are limited to those on the above list.

**Note**: The annotation tool pre-identifies discourse markers (limited to the list above). However, the tool **does not** pre-annotate discourse markers as such. Pre-identified discourse markers are displayed in red font. Annotators must **manually annotate** each of these pre-identified tokens that is acting as a discourse marker. Annotated discourse markers are displayed with red underlining. Discourse marker annotation is limited to those terms on the list above that are functioning to structure the discourse.

Many words and phrases that are used as discourse markers also have other literal meanings:

Do **you know** how many minutes we're supposed to talk for?

The situation right **now** is that we're moving in three weeks.

We will label *only* those instances that function as discourse markers.

It can sometimes be difficult to distinguish when a word or phrase is functioning as a discourse marker and when it is acting as a content word. In cases of uncertainty, annotators should leave the word or phrase unannotated.

#### 2.2.1.1 Like

One particularly difficult word that serves many functions, including acting as a DM, is *like.* It can act as a preposition, a conjunction, an adverb or adjective, a verb, and even as a substitute for "*say*" or "*said*". The following examples illustrate some of the non-discourse marker functions of *like*.

*Like* as a preposition meaning 'similar to':

They're **like** bermuda shorts but a little longer.

*Like* as a preposition meaning "as if":

It looks **like** she's wearing bermuda shorts.

*Like* used with 'to be' as a verb of quotation:

He was **like,** "I'm wearing bermuda shorts."

versus these examples, in which *like* functions as a discourse marker:

She was **like** wearing bermuda shorts.

But he played the character in it that was very **like** gross.

**Like** when I was **like** in high school and junior high school I used to hate it.

In addition to context, prosody (especially the presence of a pause) can help to distinguish cases of discourse marker from non-discourse marker *like*, although some confusing cases will remain, particularly between verb of quotation *like* and discourse marker *like*. Annotators will label only those cases in which *like* is readily identifiable as a discourse marker. This is a tricky decision to make, so the default decision is to leave *like* unannotated. Annotators should use the "difficult decision" label to register cases of uncertainty.

#### 2.2.1.2 So

*So* is another item with many different uses in discourse that may be difficult to distinguish from one another. The most common uses of *so* apart from its function as a discourse marker are as a subordinating conjunction:

We brought out pictures of her grandparents **so** she'll get to know them.

and as an adverb meaning, roughly, *therefore:*

I grew up on a farm **so** I always had outdoor pets.

As a discourse marker, *so* often serves to indicate a topic change:

I'm not sure, to be honest with you. **So,** your backpacking trip through China, that seemed to be an exciting adventure.

Discourse marker *so* can often stand alone as a complete utterance, or stand at the end of an utterance when the speaker trails off (see Section 4.4.4 for discussion of SU assignment and dangling "so"):

A: Yeah, that was a lot to go through. **So...**
B: Wow, you must be relieved to be over that.

It is often particularly difficult to distinguish uses of *so* as a conjunction versus *so* as a discourse marker. Trailing *so* as in the example above should be annotated as a discourse marker unless there exists compelling evidence that it is behaving as a conjunction (e.g., the speaker continues in the next turn with the rest of a conjoined statement). In cases of uncertainty, annotators should err on the side of exclusion, and label only those cases in which *so* is readily identifiable as a discourse marker. Cases of uncertainty should also be noted with the "difficult decision" tag.

### 2.3 Explicit Editing Terms

Explicit editing terms (EET) are fillers that occur within the context of an edit disfluency. EETs consist of an overt statement from the speaker recognizing the existence of disfluency. These typically consist of a short phrase such as *I*

*mean, sorry, excuse me, rather,* etc. Although filled pauses can sometimes function like an editing term, these will not be further annotated as EETs. Annotators should label only those cases in which a word or phrase clearly functions as an EET. Note that EETs can occur anywhere within the disfluency, including after the correction, and disfluencies can contain more than one EET. EETs are optional elements for all disfluencies. Within this document, EETs are indicated by **boldface type**.

> And when he gets free again he will have no compunction but to [complete that that same kind of lifestyle ] * **I mean** continue that same kind of lifestyle.
>
> I think [one of the positive things] * **or rather** one of the things that can come out of it is not just discipline.

## 2.4 Asides and parentheticals

Asides occur when the speaker utters a short side comment on a new topic then returns to the main topic being discussed. An aside can also occur when the speaker addresses someone who is not part of the immediate conversation (e.g., when someone else walks into the room during a phone call). Asides are often accompanied by prosodic features like a shift in intonation or the presence of a pause.

Asides can occur anywhere in the utterance. Their extent will be identified and labeled like other filler types. Within this document, asides/parentheticals (A/Ps) are indicated in **bold type** and are surrounded with {curly braces}.

> And our neighbors were, **{oh I don't know}** kind of slimy.
>
> He has now for about **{oh gosh, how long has it been}** ten years I guess.
>
> A: And I couldn't help thinking when that last question **{it was a funny question}** came up

Parentheticals are similar to asides in that they are brief remarks that break the flow of the larger utterance, but unlike asides the remark is on the same topic as the larger utterance. In standard writing, parentheticals are often accompanied by dashes or parentheses. They often display similar prosodic features to asides. As with asides, parentheticals are treated as fillers.

> The head of the United Auto Workers Union responded by calling the move **{his words}** nuts.

Parentheticals are somewhat common in broadcast news speech, while asides occur more frequently in conversational telephone speech. We do not

distinguish between asides and parentheticals for annotation purposes, but group them together as one filler type. [2]

A/Ps should be used sparingly. They should not be confused with relative clauses that modify a noun phrase (see section 4.5.4.6 for treatment of relative clauses):

> And we are presently in receipt of a site permit *which will allow us to have certain emissions up to a certain tonnage*.

Further, they are distinct from refinements and generalizers (see section 4.5.4.4) that are used to modify a previous statement:

> Right now I believe there are two policies that the government uses to modulate immigration, *three actually*.

Some very common words and phrases might be mistaken for A/Ps but should not be tagged as such. These short commentary words typically lack the prosodic features that identify A/Ps. "Say", "I think" and "for example" are common examples of non-asides that should not receive any special markup, e.g.,

> And when someone is *say* out of high school,
>
> And if he *for example* wanted to be a ballerina,

The distinguishing feature of asides and parentheticals is that they break up the stream of discourse. In standard writing, they are often represented with double dashes to set them off from the rest of the utterance. Annotators should label only those cases in which a word or phrase clearly functions as an aside or parenthetical.

## 2.5 Strings of fillers

In order to save time during annotation, long strings of contiguous filled pauses may be labeled as a single multi-word FP rather than a series of separate filled pauses.

> The **{uh, um, uh uh}** oh, what's it called, the **{uh, uh}** the Chrysler Building.

These will be separated into individual filled pause tokens as an automatic post-processing step.

---

[2] A fuller treatment of asides and parentheticals would distinguish the two types, and would not classify them as fillers since they contain material that should be preserved during cleanup. However, for ease of annotation, they are classified as fillers under this version of SimpleMDE and no distinction is made between the two types. This filler type is not evaluated. The treatment of asides and parentheticals will be revisited in future task definitions.

Because they typically consist of multiple words, strings of multiple DMs or EETs must be annotated as *separate* units rather than as one multi-token unit:

**{Like}, {you see},** that's a hard question.

***NOT***

**{Like, you see},** that's a hard question.

Fillers of different *types* that occur in sequence should be annotated separately according to their type. For instance,

**{Um, uh} {well} {you see} {uh}** it's not that simple.
*Annotation:* **FPause Disc.Marker FPause**

Upon post-processing, this example would be rendered as follows:

**{Um} {uh} {well} {you see} {uh}** it's not that simple.
*Post-process:* **FP FP DM DM FP**

### 2.6 Interaction among filler types

Occasionally a filled pause or discourse marker can occur within the extent of another filler. When this occurs, annotators should separately label the FP or DM as its own filler type, within the extent of the larger filler. For instance, consider this example of a discourse marker occurring within an aside/parenthetical (DM *like* is underlined here for clarity):

I talked about how a lot of the problems they have to overcome **uh** {**It's a very like complex situation**}, to go into space.

Recall that two or more fillers of different types that occur next to one another should be annotated as separate fillers. (See Section 2.1.1 for further discussion on this point). Further recall that FPs functioning as EETs should not be further annotated as EETs but will be labeled only as FPs (see Section 2.3).

## 3 Edit disfluencies

### 3.1 Introduction

Edit disfluencies are portions of speech in which a speaker's utterance is not complete and fluent; instead the speaker corrects or alters the utterance, or abandons it entirely and starts over.

### 3.2 Structure of edit disfluencies

Edit disfluencies have a more complex internal structure than fillers, consisting of the depod, interruption point and correction, each defined below. For SimpleMDE annotation, annotators will identify the extent of the depod and interruption points. The correction is not explicitly labeled. However, a brief explanation of each component will help annotators recognize disfluencies and tag them correctly.

- **Depod**
  The depod is the **de**letable **po**rtion of the **d**isfluency. It is also known as the original portion of the utterance. This is the speaker's initial attempt that exhibits some kind of disfluency and later gets corrected or (in the case of restarts) abandoned entirely. The entire depod is discarded upon cleanup of the transcript. All edit disfluencies have a depod component.

- **Interruption point**
  The interruption point or IP is the point at which the speaker breaks off the depod with an editing term, a repetition, revision or restart. All edit disfluencies have at least one interruption point at the right edge of the depod. Complex disfluencies contain multiple interruption points.

- **Correction**
  The correction consists of the portion of the utterance that has been repaired by the speaker and is fluent. The correction is the part of the edit disfluency that remains after cleanup of the transcript. The correction will be empty for restarts.

### 3.3 Annotation of edit disfluencies

**Depod:** Annotators select the full extent of the deletable portion for each disfluency, and apply the depod label to those words (shown in [square brackets] within this document).

**Interruption point:** Interruption points at the right edge of a depod are automatically identified and labeled by the annotation tool. Annotators will not explicitly label such IPs. Multiple IPs within the extent of a complex disfluency depod will be manually labeled by annotators. (IPs are indicated by asterisk * within this document.)

**Correction:** For SimpleMDE, annotators do not explicitly tag the corrected portion of edit disfluencies. (Corrections are underlined within this document).

**ED type:** Within SimpleMDE, annotators do not classify edit disfluencies according to their type. (Note that this means annotators do not distinguish simple edits from complex edits.)

### 3.4 Types of edit disfluencies

Edit disfluencies can be divided into four categories: *repetitions*, in which the speaker repeats the same word, part of a word or sequence of words more than once; *revisions*, in which the speaker explicitly modifies and replaces a portion of the utterance; *restarts*, in which the speaker abandons the utterance s/he's begun and simply starts over; and *complex disfluencies*, which consist of multiple or nested edits.

For purposes of SimpleMDE, annotators will **not** distinguish between the various types of edit disfluencies. However, for explanatory purposes, an inventory of ED types is presented in the sections below.

### 3.4.1 Repetitions

In the case of repetitions, the depod is repeated in the corrected portion of the utterance. The repeated portion can be a word fragment, a complete word, or more than one word:

> [He-] * he's really out of line, or at least that's what I was told
>
> [The thing] * the thing I wanted to cover with you is
>
> [After the] * **um** after the third paragraph, what did it say next?

In cases of repetition, cleanup (removal) of the depod does not alter the meaning of the utterance at all.

Keep in mind that not all repetitions are disfluencies. Examples like repeated backchannels and the use of repetition for emphasis are entirely fluent and should not be labeled as EDs:

> We had a really, really wonderful time.

### 3.4.2 Revisions

Revisions are very similar to repetitions, but the corrected portion that replaces the depod modifies its meaning, rather than simply repeating it in whole or in part. The revision might change the meaning entirely, or may modify or clarify it only slightly. For instance,

> Our children like [the dog] * **I mean** the cat, that the neighbors got.
>
> Show me flights [from Boston on] * **uh** from Denver on Monday.
>
> Three hundred fifty-six residents were [killed] * **er** injured **rather.**
>
> So we built a cradle for it, and [we got th-] * once it was turned we got one cutout on the saw.
>
> [Do you] * by mistakes do you mean just like honest...

#### 3.4.2.1 Revisions with information loss

Another complication that occasionally affects revision disfluencies occurs when information is present in the depod, but that information is not contained within the revision. This most often happens when the depod contains a full noun or noun phrase, but the corrected portion contains only a pronoun:

I gave a book [from Susan] * **I mean** by her to Ralph.
I flew [to Dallas] * **I mean** from there, yesterday.

Revision disfluencies of this type are very unusual, and they will be annotated in the standard way, despite the fact that this will result in a small loss of information in the cleaned-up utterance.

### 3.4.3 Restarts

In the case of restarts, also known as false starts, a speaker abandons an utterance or constituent, and neither corrects it nor repeats it partially or wholly, but instead restructures the utterance and starts over. A restart is often followed by a noticeable pause and may be accompanied by other prosodic features like resetting of pitch. In restarts, the corrected portion is empty. This means that important information contained within the restart depod may not be repeated or restated by the speaker later in the discourse.

Most typically, a restart does not add information to the overall discourse. These are cases in which the speaker abandons an utterance or constituent and then starts over, restructuring the original utterance entirely:

[Do you know whether he wants] * how does he like it to be cooked?

I can't really say [that there should be a] * what type punishment there should be.

You know, [they always end up] * sometimes they change their minds.

[It's also] * I used to live in Georgia.

Sometimes, however, the original utterance contains important information content that may not be presented elsewhere within the discourse:

A: [I happen to live not too far away from] * well, I've actually worked for the company that has been blamed for the Challenger disaster.
B: Oh, is that right?

In order to enhance inter-annotator consistency and maximize efficiency, all restarts are labeled as depods, even when they contain additional information that may not be present elsewhere within the discourse.

Note that restarts are easily confused with incomplete SUs. Restarts are limited to those cases where a speaker interrupts him or herself, restructuring the utterance and continuing with the discourse. See Section 4.4.4.1 for further discussion.

### 3.4.4 Complex disfluencies

In spontaneous speech, speakers often produce a series of disfluencies in succession. These can be serial, where one occurs directly after another, or they can be nested, where some component of one disfluency itself contains another disfluency.

#### 3.4.4.1 Complex original utterance

In cases where the original utterance itself contains multiple (serial, adjacent) disfluent utterances, the *maximal extent* of the disfluent portion will be taken as the depod. To indicate that the depod contains additional internal structure, annotators will explicitly identify the interruption points contained within the depod. The final interruption point at the right edge of the depod is automatically identified for all disfluencies and is not separately annotated.

```
Yeah but [the * the big * the b- * the big] * um the betrayal or
whatever she called it

[It's * this is like only like the third or fourth time I've * I
ne- * I'm real bad about] * I never make the phone calls
```

#### 3.4.4.2 Complex correction

In some cases, the corrected portion of one edit disfluency contains another disfluency:

```
                      [---corrected--]
1)    I'm sure [the] * that  [the]  the staff learn what's normal.
                       [disf]

                            [---corrected ---]
2)    I flew [from Dallas] * uh from [Aus-] Austin this morning.
                                  [disf]

                                          [--corrected---]
3)    It gets kind of expensive [to renew] * to [renew] renew it.
                                             [disf ]
```

Annotating overlapping structure within a complex disfluency is beyond the scope of SimpleMDE annotation. Instead, annotators will identify just the depod section of each disfluency. As a rule of thumb in these cases, annotators should take the rightmost (final) instance of each word within the disfluency as the corrected portion; the other instances of each word are part of the depod. For SimpleMDE, the examples above are annotated as follows:

```
1)    I'm sure [the] * that  [the] *  the staff learn what's normal.
               depod         depod

2)    I flew [from Dallas] * uh from  [Aus-] * Austin this morning.
              ---depod---     FP       depod

3)    It gets kind of expensive [to renew] * to [renew] * renew it.
                                --depod--        depod
```

#### 3.4.4.3 Complex disfluency and restarts

In many cases a complex disfluency ends in a restart, or the restart will contain another disfluency within it, e.g.:

> [I hoped that we could be better than * than **um** expecting], **Well,** we should have higher expectations than that.

The repetition contained within this restart (than than) cannot be separately annotated as a repetition because of the prohibition on nested disfluencies. In cases like these, annotators should select the maximal extent of the depod – in this example, the full extent of the restart depod – and should indicate any internal interruption points within that extent.[3]

#### 3.4.4.4 Multiple disfluencies in one utterance

When an utterance contains multiple or serial disfluencies, the annotator should apply the rules for annotating edits that are described above: selecting the maximal extent of each depod; annotating internal interruption points; and treating the rightmost occurrence of each word within the disfluency as the correction.

> [Did they, * did they] like on bottles, did they give you [a] so many cents back?
>
> [When I] when I see that money taken out of my paycheck [each * each week * **or** each] every other week, [I] I really think that money's history.
>
> Yeah, [I've * I've heard some] **I mean** I've heard [sta- * sta-] statistics and things.

### 3.5 Disfluencies across turn boundaries

In some cases, disfluencies can occur across same-speaker turn boundaries:

> A: [I ha-] *
> A: I hated that car so
> A: it burnt up and I did not set it on fire.
> A: I have been accused of that but I did not.

These cases should be annotated in the same way as within-turn disfluencies.

### 3.6 Disfluencies and conjunctions

In many cases, a conjunction will occur at the edge of an edit disfluency. The conjunction should be excluded from the span of the depod unless the conjunction itself is disfluent (that is, unless the speaker either repeats, revises or

[3] Note again that the SimpleMDE task is limited to identification of the maximal extent of depods plus any internal interruption points in depods with multiple IPs. This annotation task does not map directly onto the identification of complex edits, nor does it support identification of the extent of the original portion of an edit disfluency in the case of nested edits. The issue of complex edits will be revisited in future task definitions.

restructures the conjunction in the corrected portion of the disfluency). For example, the conjunctions "but" and "because" (italicized for emphasis) are excluded from the depod extent in the following examples:

I agree, *but* [I] * I feel that **uh** a lot of people have gotten lazy about voting

*because* [that was]* The person meant to say this.

When the conjunction itself is disfluent, it must be contained within the depod, as in these examples:

[and if] * and when he gets free again he will continue that same kind of lifestyle

[and i-] * but if he's not in the mood you'd better stay away from him

## 3.7 Disfluencies and fillers

### 3.7.1 Depods within the extent of a filler

Depods that occur within the extent of a filler (for instance, a repetition within an aside) should be treated as a regular depod.

### 3.7.2 Fillers within the extent of a depod

Fillers that occur within the context of a depod are to be annotated separately as fillers, but are contained within the extent of the depod. For instance,

[I think **uh** that] * I think that is a tough one.

### 3.7.3 Fillers at the edge of a depod

Fillers that occur at the right or left edge of a depod should not be included within the extent of the depod, but should be separately annotated as fillers. For instance,

**Um** [but the job that I ju-] * I had this job that I lost.

[But to just] * **Well** I don't know exactly.

### 3.7.4 Fillers and complex edits

When a filler occurs between two adjacent depods, the filler is excluded from the span of either depod. Instead, the two disfluencies should be indicated as two simple depods, with a filler standing between them:

[I] * **uh** [I don't] * I don't have any cats.

**NOT** [I * **uh** I don't] * I don't have any cats.

However, if a filler does not occur between two disfluencies, but inside one of them, the filler should be included within the maximal depod extent:

```
[I * I really uh haven't] * I don't have any cats.
```

### 3.8 Disfluency conclusions

In labeling difficult examples, annotators should keep in mind that the goal of metadata annotation is to create more readable transcripts while preserving the meaning of the discourse. Annotators should not just read the transcript, but also listen to the speech, and should use their best judgment about the speaker's intended meaning when determining how to properly label edit disfluencies.

## 4 SUs

### 4.1 Introduction

One of the goals of metadata annotation is the identification of all units within the discourse that function to express a complete thought or idea on the part of the speaker. Often times this unit corresponds to a sentence, but sometimes a sentence will contain multiple units; other times a unit is semantically complete but smaller than a sentence. For the purposes of metadata annotation, these elements are called SUs. We will not take a stand on the issue of what SU stands for, but some possibilities include: Sentential Units, Syntactic Units, Semantic Units and Slash Units. As with disfluency annotation, the goal of SU labeling is to improve transcript readability by creating a transcript in which information is presented in small, structured, coherent chunks rather than long turns or stories.

Every word within the discourse must be assigned to an SU, and all SUs must be classified according to their function within the discourse. Although SUs typically occur within a single speaker turn, they can occasionally span a turn, particularly in the case of failed interruptions and overlapping speech. SUs can be whole well-formed sentences, phrases or single words. Many short SUs that are not full sentences or clauses but are nevertheless complete units serve to regulate the discourse.

In identifying SUs, annotators should rely on both the transcript and the audio. Syntactic and prosodic factors can influence the placement of an SU boundary. In read speech, SUs are usually easy to identify and are often accompanied by a pause. However, breath groups may not always correspond to SU boundaries in read speech; sometimes a breath occurs in the middle of an SU. Spontaneous speech presents a more difficult problem, since pauses do not always occur at major constituent boundaries, and other metadata phenomena like fillers and disfluencies can further obscure the presence of SUs. In all cases, annotators should rely primarily on semantic and syntactic information, and secondarily on prosodic information, to inform their decision about where to place the boundary.

The following sections provide guidelines for recognizing the existence of SU boundaries and categorizing them according to their type. Notice that we refer to the SU *boundary* when we are discussing the annotation task. We do so because the annotation task is centered on the detection and classification of the breaks *between* the SUs in a discourse.

During annotation, annotators label only the final word of an SU; the next word uttered by the same speaker will automatically be interpreted as the beginning of a new unit. Every word in the discourse must be assigned to some SU. Each word contained between two SU boundaries is considered *part of* the same SU. Annotators will distinguish two kinds of SUs: sentence-level and sentence-internal.

### 4.2 Background information for non-linguists

Although deciding between SU types is relatively straightforward, it can be difficult to determine when to begin a new SU boundary and when to place two segments within the same SU. In order to enhance inter-annotator consistency, we will define SUs on the basis of some easily recognized basic syntactic units. While some sentences constitute single and complete SUs, many sentences contain multiple SUs; moreover, in conversational speech, sentences are not always well-formed or easily detected.

Clauses, on the other hand, are relatively easy to identify. A clause can be thought of as something like an 'atomic sentence.' It is merely a subject plus a single simple predicate. For example, the sentence:

```
If the annotator understands clauses then she sees them both.
```

Consists of the two clauses:

```
the annotator understands clauses
she sees them both
```

We will depend a great deal on the notion of a *clause* in the discussion that follows. For both sentence-internal and sentence-level SU boundaries, the ideal placement of a break is between clauses. There will always be an SU break at the boundary between clauses; the only remaining question, then, is what type of clause boundary we are dealing with.

The data being annotated for SimpleMDE is quite slippery with regard to the placement of SU boundaries, so it will be important to distinguish between a few parts of clauses as well. Specifically, the sections that follow will make use of the terms clause, predicate and noun phrases in the manner illustrated by the following set of examples:

Clauses (S):

```
Sandhya loves annotation projects.
Ramez went to the EARS workshop.
Meghan is typing on her laptop.
```

Predicates/Verb phrases (VP):

```
loves annotation projects
went to the EARS workshop
is typing on her laptop
```

Noun phrases/subjects (NP):

```
Sandhya
Ramez
Meghan
the EARS workshop
her laptop
annotation projects
```

While the examples in this section are very straightforward and not representative of the kinds of data to be annotated for SimpleMDE, the concepts they illustrate will apply to the whole range of data that annotators will encounter during SimpleMDE annotation.

The following sections provide more detailed guidelines for recognizing the existence of SUs and categorizing them according to their type. The examples that follow are all taken from real conversational and broadcast news speech data.

### 4.3 Sentence-level vs. sentence-internal SUs

Distinguishing sentence-internal SUs from those that occur at sentence boundaries is an important component of the SU tagging task. Sentence-level SU breaks are fundamental and directly support the SU research task. Sentence-internal breaks are secondary and have been introduced to support inter-annotator consistency.

Sentence-level SU breaks are used to indicate the presence of a main (independent) clause. These independent main clauses can stand alone as a sentence and do not depend directly on the surrounding clauses for their meaning; that is, the proposition asserted by this main clause does not depend on another clause for completion.

Sentence-level SU breaks are inserted between sentence boundaries; within each sentence further internal breaks might also exist. In most cases, sentence-level SUs would be represented in standard writing with end-of-sentence punctuation like periods or question marks. Sentence-level SUs fall into one of four categories: Statements, Questions, Backchannels and Incomplete-SUs. The four SU categories are described in detail in Section 4.4.

Sentence-internal SU breaks signal units that are smaller than a main clause and cannot stand alone as a complete sentence[4]. These smaller units are typically non-main clauses; they may also be units that are smaller than a clause but still constitute an SU. In many cases, sentence-internal SU breaks would be represented in standard writing with a comma. In all cases, these smaller units cannot stand alone as complete sentence-level SUs because they are conditioned by adjacent clauses that affect their meaning. Sentence-internal SUs can also be used to indicate boundaries between larger clauses containing both subject and predicate, when these clauses depend on one another for their full meaning. A clear example of this is an if-then construction:

```
If you never even see the money, then you don't spend it.
```

We can easily detect the two separate clauses that make up this sentence:

```
you never even see the money
you don't spend it
```

Each of these clauses constitutes a main clause, containing both subject and predicate. However, neither of these clauses can stand on its own and still retain the meaning expressed by the full sentence. The proposition expressed by this sentence relies on both clauses interacting with one another. Therefore, we would not want to draw a sentence-level SU boundary between the two clauses, but we still want to capture the fact that we clearly recognize two separate clauses in this sentence. We rely on a sentence-internal SU boundary to do this.

Sentence-internal SU breaks fall into two categories. Coordination SU breaks identify two clauses which are not full sentences but which are joined by a coordinating conjunction. Clausal SU breaks identify non-sentence clauses that are joined by subordination and a handful of other phenomena. Specific rules for recognizing and classifying sentence-internal SU breaks are described in Section 4.5.

The goal of SimpleMDE SU annotation is for annotators to recognize every clause in the discourse and to impose upon it the appropriate SU label. By forcing a decision at each clause boundary, we eliminate the more difficult task of identifying both the extent and the type of each discourse-level (sentence-external) SU in isolation, a task that has proven impossible to execute with any reasonable degree of efficiency or consistency. A handful of additional rules are necessary to cover common features in spoken language that don't relate to clause boundaries; these rules are fully described in the sections that follow. For the vast majority of cases, however, the SU annotation task involves recognizing and categorizing clause boundaries.

[4] Sentence-internal SU breaks are introduced solely for ease of annotation within the current definition of SimpleMDE. Identification of sentence-internal SUs is not an MDE research task.

When deciding between the two types of SU breaks, annotators should look first for evidence of a sentence-level break. Annotators should treat the sentence-level SU as the default break type, and should impose a sentence-internal SU only when no evidence of a sentence-level break exists.

## 4.4 Sentence-level SUs

### 4.4.1 Statement

The statement label is used for a complete SU that functions as a declarative statement. The expected end-of-sentence punctuation for a statement is a period (or exclamation point).

```
I thought it was a strange topic about corruption /.
```

### 4.4.2 Question

The question label should be used for a complete SU that functions as an interrogative. The expected end-of-sentence punctuation for a question is a question mark; but keep in mind that SUs may be smaller than a sentence.

```
How many hours a week are you supposed to work /?
```

#### 4.4.2.1 Tag questions

A tag question is a phrase added to the end of an utterance that invites the listener to give feedback. Tag questions should be labeled as questions**,** and should be separated from the previous SU:

```
A: It seems like winter will never end /. doesn't it /?

A: You've been working there for years /. haven't you /?
```

#### 4.4.2.2 Question intonation in statements

Some speakers utter statements with rising final intonation so that they sound like questions. When the utterance is clearly functioning as a statement, it should be tagged as such. When the utterance is clearly asking a question, it should be tagged as a question. In cases of uncertainty, annotators should categorize the SU as a statement, and add the "difficult decision" label.

### 4.4.3 Backchannel

A backchannel, sometimes called an acknowledgement or continuer, is a word or phrase that provides feedback to the dominant speaker, indicating that the non-dominant speaker is still engaged in the conversation. Backchannels serve a function similar to gestures like nodding one's head. For instance,

```
A: You know, it's just been really difficult for me /.
B: Uh-huh /@

A:  If it happens again I'm going to have to say something /.
B:  Yeah, yeah /@
```

Backchannels can sometimes preface longer responses to questions or statements by the other speaker, but those direct responses themselves are not backchannels (they are statements):

```
Oh /@ I didn't realize that's what you were talking about /.
```

Because it is sometimes difficult to recognize and accurately label backchannels, and because the list of potential backchannels is very large, annotators will tag only a subset of all possible backchannels during SU annotation, limited to the following items:

| | |
|---|---|
| hm/hmm | right |
| huh | sure |
| mm-hm/mm-hmm | yeah, yea |
| oh | yep |
| okay, OK | yes |
| really | uh-huh |

Some words not contained on this list may function as backchannels, and the list may be expanded in future task definitions, but for this version of SimpleMDE backchannels are limited to those on the above list. Annotators may encounter alternate spellings of the backchannels on this list. When it is clear that the alternate forms correspond to the words on this list, they should be tagged as regular backchannels.

Keep in mind that many of these words may be used in a non-backchannel way. For instance, many of these words may serve as a direct response to a question, in which case they should be annotated as a statement. (Note that the word "no" typically serves as a question response, and unlike "yeah" it never functions as a backchannel.) Similarly, the potential backchannel may be used as a question (e.g., "huh?"). Annotators should label only those cases in which these words are functioning in a way that is clearly recognizable as a backchannel.

**Note**: The annotation tool pre-identifies potential backchannels (limited to the list above). However, the tool **does not** pre-annotate backchannel SUs as such. Pre-identified backchannels are displayed in green font. Annotators must **manually annotate** each of these pre-identified tokens that is acting as a backchannel SU. Annotated backchannel SUs are displayed with /@. Backchannel SU annotation is limited to those terms on the list above that are unambiguously functioning to as backchannel responses.

#### 4.4.3.1 A note on "yeah"

The word "yeah" can take on several different roles in the discourse. As a rule, when "yeah" occurs at the beginning of a turn and is not a response to a question, it will be considered a backchannel.

However, when a speaker uses "yeah" in the middle of an utterance or a turn, as a prefix to the following complete SU, it is not separately annotated. For example:

**Backchannel**

**Yeah /@** it's like **/,** I'm young **/&** and I should get started getting into shape **/.**

**Response**

A: You said you had ten cats **/?**
B: **Yeah /.** When I moved **/,** I gave them to my mother **/.**

A: Did you say you're calling from Dallas **/?**
B: **Yeah /.**

**Not annotated**

Somebody can look **/&** and see **/,** yeah ∅ you laid fifteen thousand bricks today **/.**

Of course my family has always had cars like that **/,** in that yeah ∅ my father has always had a VW bus **/.**

#### 4.4.3.2 Backchannels at the start of a file

Conversations cannot begin with a backchannel, because backchannels by definition provide feedback to the dominant speaker. As a convention, when the first word in a transcript includes any of the potential backchannel words, even if the annotator suspects this might be a backchannel to a statement not included in the transcript, it is not treated as a backchannel but rather as a Statement SU, e.g.,

Okay **/.** Carol **/.**

All right **/.** Family is the topic**/.**

#### 4.4.3.3 Backchannels within a longer utterance

When a backchannel begins or ends a longer utterance, it is usually labeled as a separate SU.

A: Right **/@** so you just bought another Plymouth then **/?**

Sometimes speakers employ potential backchannel words as a sentence prefix in the middle of a turn; in these cases, the token functions something like a discourse marker:

A: So how do you make this soup **/?**
B: First you take a couple of carrots and chop them **/.** Okay ∅ and then you sauté them in butter **/.**

In these cases, the token should **not** be specially tagged but should be included within the extent of the larger SU, because it is not functioning as a backchannel.

When a potential backchannel word occurs in the *middle* of a longer SU, it should be contained within that SU and should not receive a separate backchannel SU tag, e.g.,

```
My family has always had cars like that /, in that yeah ∅ my
father has always had a VW bus ever since about nineteen sixty /.
```

When the first word in a speaker turn includes a potential backchannel, it can be tagged as either a statement or a backchannel, depending on its function. When the word serves as a response to a question, it should be classified as a statement:

```
A: How are you doing /?
B: Okay /.
```

If the word provides feedback to the dominant speaker, it is a backchannel:

```
A: I've lived in Friendship Heights for years /.
B: Okay /@
```

When in doubt, the word should be tagged as a statement, not a backchannel (according to the SU hierarchy, Section 4.6).

#### 4.4.3.4 Chains of backchannels

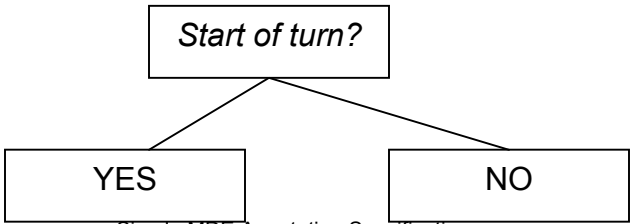
In many cases a speaker will chain together several backchannels. To enhance inter-annotator consistency and increase efficiency in these cases, annotators should create a single SU boundary for the entire chain of backchannels, regardless of the presence or absence of pause, breath, change in intonation or other prosodic features.

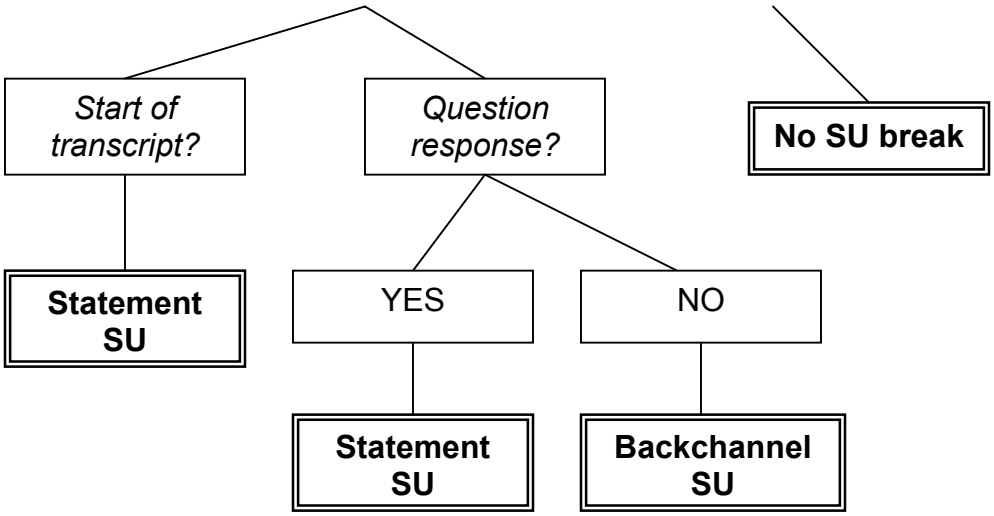
```
A:  Hmm, yeah, okay /@
B:  Yeah, right/@ I see what you mean /.
```

These will be separated into individual backchannel SUs as an automatic post-processing step. When chains of backchannels cannot span a speaker turn but must be separately annotated.

#### 4.4.3.5 A backchannel decision tree

The following decision tree can help the annotator distinguish among the roles for potential backchannel words:






### 4.4.4 Incomplete SUs

When an utterance does not constitute a grammatically complete sentence, phrase or continuer, and does not express a complete thought, it is labeled as an incomplete SU. Incomplete SUs occur in two situations. When a speaker trails off and completely abandons an utterance, without restructuring it and continuing along the same lines, an incomplete SU exists. In standard writing, the utterance might be followed by ellipses (…). For instance:

```
A: If you put enough patience into <pause> /-
B: <pause> Yeah /@ just be consistent and diligent /.
A: Right /@
```

The other type of incomplete SU occurs when one speaker is interrupted by another speaker:

```
A: Yeah /@ but the thing about /-
B: No, see, you have to take inflation into account /.
A: Oh, okay /@ I get it now /.
```

#### 4.4.4.1 Incomplete SUs and Restarts

Incomplete SUs are easily confused with restart disfluencies (see Section 3.4.3). For purposes of SimpleMDE annotation, restarts occur when a speaker interrupts him/herself and then restructures the utterance and continues speaking on the same topic, e.g.:

```
A: [Do you know] * How would he like it to be cooked /?
```

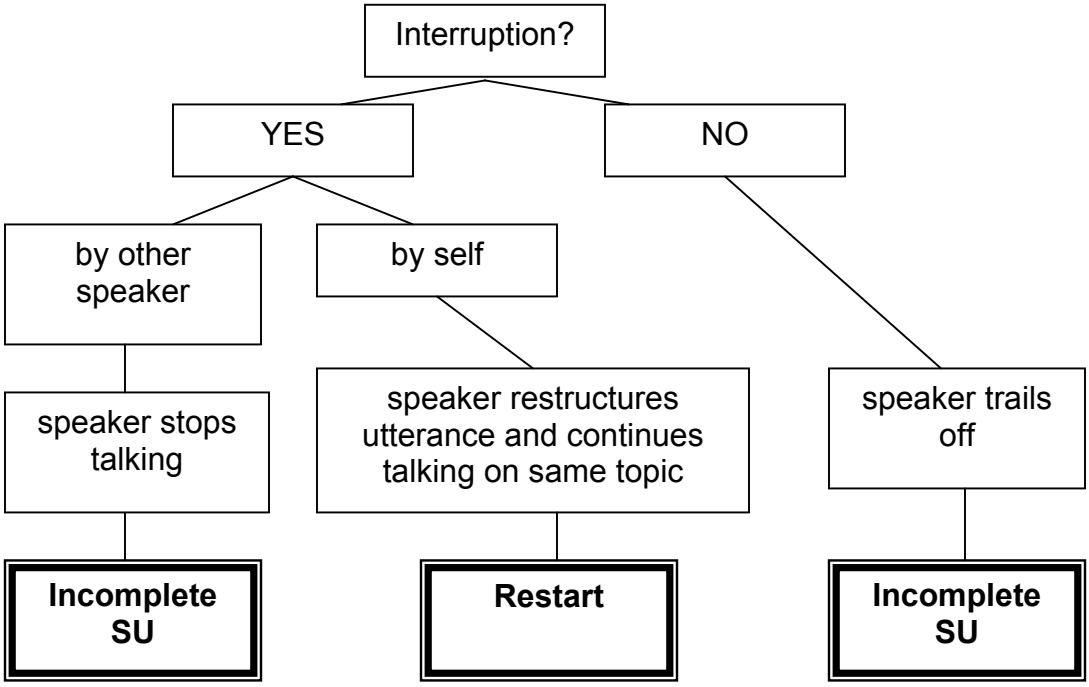
Restarts are labeled as depods, and are contained within the larger SU. Restarts do not induce a separate SU break of their own.

Incomplete SUs, on the other hand, occur when a speaker is interrupted or when the speaker trails off, failing to complete the utterance within a turn:

```
A: Yeah /@  but the thing about /-
B: No, see, you have to take inflation into account /.
A: Oh /@ okay /@ I get it now /.
```

Incomplete SUs generate a sentence-level SU break.

A simple decision tree can assist the annotator in deciding between Incomplete SU and Restart:




### 4.5 Recognizing SU breaks

Although deciding between SU types is relatively straightforward, it can be difficult to determine when to begin a new SU boundary and when to place two units within the same SU. The following sections provide a number of guidelines for recognizing SU boundaries for all SU types.

#### 4.5.1 Phrases

Short phrases that do not constitute grammatically complete sentences but nevertheless function as a complete sentence (as in a direct response to a question) receive the statement SU break:

```
A: What did you get for your birthday /?
B: A bunch of CDs /.
```

```
A: Cool /.
```

### 4.5.2 Simple sentences

An SU can comprise a simple sentence with a single main clause:

```
The last great explorer Jacques Cousteau has died in Paris at age
eighty-seven /.
```

### 4.5.3 Compound and complex sentences

Compound sentences consist of two or more clauses joined by coordination, while complex sentences contain some kind of subordination. Compound and complex sentences present a significant challenge to the annotation process, but they are very common in discourse. The number of SUs present in a given complex or compound sentence will vary depending upon both the number and type of clauses contained within the sentence.

#### 4.5.3.1 Coordination

Compound sentences involve some kind of coordination; that is, they contain multiple phrases or clauses joined by a coordinating conjunction (CC). We will consider correlative conjunctions (conjunctions used in pairs) as part of this set.

The class of CCs is semi-closed: there may be new elements, but these are limited in number and frequency and typically amount to synonyms of current members. Moreover, coordination can occur even without the presence of an explicit coordinating conjunction. A list of some of the commonest CCs follows. Note that this list should be considered illustrative, rather than exhaustive.

```
--NULL--                              yet
and, and so, and yet                  both… and…
but, but also, but only               not only… but also…
for                                   not because… but because…
nor                                   either… or…
or                                    neither… nor…
so
```

##### 4.5.3.1.1 Coordination of main clauses

Coordination can occur between two or more main clauses. This is the simplest type of compound sentence. A sentence containing two or more main clauses joined by a coordinating conjunction should be divided with a sentence-level SU break before the conjunction:

```
Most folks on this planet spend virtually all of their lives on
dry land /. but Cousteau was rarely found out of the water /.
```

Each main clause element on either side of the coordinating conjunction is part of a separate SU, with the division point coming before the coordinating conjunction. Two separate sentences joined with a coordinating conjunction are treated in the same way:

```
We went to the ski show /& and looked at all of the resort
information /.
And we picked up a lot of brochures /.
```

#### 4.5.3.1.1.1 Main clauses with semantic dependency

When the coordinating or correlating conjunction joins two main clauses that depend on one another for completion of an idea, a sentence-internal coordination SU break is imposed.

```
Not only do we methodically destroy the coastal fringe /& but we
also throw back our toxic effluence directly in the sea /.

Republicans argued the climate did not create President Clinton's
problems /& but he did /.
```

The clauses on either side of the conjunction cannot stand alone as separate sentences, but must remain linked to one another in order to express a complete idea. Therefore, we separate the clauses with a coordination SU and preserve a single sentence.

#### 4.5.3.1.1.2 Main clauses without a conjunction

When a sentence contains multiple main clauses that are coordinated without conjunctions, they should be separated into individual sentence-level SUs of the appropriate type. Each unit expresses a complete idea.

```
NATO defense ministers in Brussels yesterday issued the toughest
warning yet to Yugoslav president Slobodan Milosevic /. Stop the
killing of ethnic Albanians who live in Kosovo /. Withdraw your
heavy forces /. And begin peace talks /. or face the possibility
of military action by the west /.
```

These clauses might be separated by additional punctuation (dashes, commas, semicolons, colons) in standard writing, but punctuation is highly variable and should not be the deciding factor in determining where to place SU boundaries.

#### 4.5.3.1.2 Coordination of non-main clauses

Coordinating conjunctions can also join two or more non-main clauses. Subject and predicate clauses are not main clauses; that is, they cannot stand alone as a complete sentence. Coordination between non-main clauses imposes a sentence-internal coordination SU break.

For instance, consider this case of predicate coordination:

```
You can't stand in front of a counter /& and do that kind of
stuff all day /& and feel like you're contributing something to
scientific knowledge /.
```

or this case of subject coordination:

```
Also vacation time too is a big one for me /& and support for
whatever other kinds of professional activities you have at work
/.
```

or this case of prepositional phrase coordination:

```
And then I've worked with regular industry type places /& and
then with others that are kind of half way between /.
```

For non-main clause coordination, the clauses remain within the same sentence and receive only one end-of-sentence SU boundary, but the clauses are also separated by a coordination SU break. The following examples illustrate the difference between coordination of main clauses, resulting in multiple sentence-level SU breaks:

```
A: I had to cut it into slabs /. And I had to build up this big
square /. And then I turned it on the lathe /.
```

versus coordination between non-main clauses, resulting in multiple coordination SU breaks but only one sentence-level break:

```
A: I had to cut it into slabs /& and build up this big square /&
and turn it on the lathe /.
```

#### 4.5.3.1.3 Conjoined simple subjects

When multiple subjects or other NPs are conjoined with a single predicate, they do not receive a SU break:

```
It's like being a painter ∅ or an artist /.

If American young men ∅ and women are in harm's way

John ∅ and Mary Smith …

I now have a letter from Trent Lott ∅ and other Senate leaders /.
```

#### 4.5.3.1.4 Lists

Strings of NPs or predicates that function as a list should not be separated by SU breaks:

```
And I take Discover Magazine ∅ and also CD and Stereo Review ∅
and National Geographic I think ∅ and Boys Life /.
```

### 4.5.3.2 Subordination

Subordination turns a clause into something that depends on the rest of the sentence for its meaning. Subordinate clauses cannot stand on their own as complete sentences; therefore, they cannot by themselves constitute a sentence-level SU.

When the clause begins with a subordinating word, it is no longer an independent clause; it is a dependent or subordinate clause because it depends on something else (the independent clause) for its meaning. Consider the following sentence:

```
If the conflict cannot be resolved, European Champion Germany
could replace France.
```

In this sentence, *if the conflict cannot be resolved* is the dependent or subordinate clause; *European Champion Germany could replace France* is the independent clause. The two clauses are semantically linked: they depend on one another for meaning. More specifically, we can say that the subordinate clause is semantically dependent upon the independent clause.

Subordination is an important concept for SU annotation in that two clauses that have a subordination relationship cannot be separated into two distinct sentences, because they depend on one another for meaning. Instead of creating a sentence-level SU break, subordination creates a sentence-internal clausal SU break.

#### 4.5.3.2.1 Types of subordination

##### 4.5.3.2.1.1 Subordinating conjunctions

Most subordination relationships are created by subordinating conjunctions. A large number of words and phrases can act as subordinating conjunctions. A non-exhaustive list of some of the most common subordinating conjunctions follows:

| | | |
|---|---|---|
| after | even though | though |
| although | if | unless |
| as | rather than | until |
| because | since | when |
| before | so that | while |

The following sentences contain clauses joined by subordinating conjunctions. For clarity's sake, the conjunctions appear in *italics.*

```
It's an opportunity to stay alive really /, instead of trying to
exist on the good health of my youth /.

And now that El Nino is virtually gone /, there is La Nina to
worry about /.

While mourners gather for his funeral /, colleagues and friends
remember his contributions to our understanding of the last
frontier /.

because there is not enough to go around for everybody /, you
don't expect anything /.

He'd probably look funny /, if we did that /.
```

Legally speaking, the sanctions cannot be lifted /, *until* the special commission issues a positive report to the security council /.

And they will not have Reeves on the sidelines /, *when* they meet Detroit on Sunday /.

*As* we continue with this special edition of NBC Nightly News /, we want to go to Charleston, West Virginia, where former Secretary of State Henry Kissinger has been making an appearance this evening /.

Note that the order of the clauses may be inverted so that the conjunction comes at the beginning of the sentence rather than between the two clauses. In any case of subordination due to a subordinating conjunction, annotators should tag the boundary between the clauses with a clausal SU.

#### 4.5.3.2.1.2 Subordinating verbs

Subordination can also be caused by structures that do not include a subordinating conjunction, as in the case of certain verbs. These verbs often express thinking, feeling, saying and the like, and may often be followed by a complementizer, like “that” or “whether”. A non-exhaustive list of verbs that trigger subordination follows:

*verbs of perception and discovery*
find, hear, notice, realize, recognize, see, detect, sense

*verbs of thinking and feeling*
believe, decide, feel, figure, hope, know, prefer, remember, seem, suppose, suspect, think, wish, assume, appreciate

*verbs of declaration*
deny, discuss, explain, say, mean, tell, claim, mention, articulate

*verbs of questioning*
ask, guess, wonder, question

In these cases, a clausal SU break should be placed after the verb and before the complementizer, if present. For instance,

A friend of mine at work here said /, that he tried it with his dog /.

Note that when such verbs occur medially within a larger SU do not invoke subordination and therefore do not trigger a clausal SU break:

That would I think ∅ be a problem.

#### 4.5.3.2.1.3 Conditionals

Another common cause of subordination are conditionals. If-then statements are the most common type of conditional, e.g.

```
If someone were studying economics /, then the State Department
could offer to send them to a foreign country for two years /&
and teach them how to run the country /.

The Jags will clinch the AFC Central title /, if the Tennessee
Oilers lose or tie the Green Bay Packers /.
```

Note that not all conditionals follow this exact structure. The order of the clauses is not fixed, the “then” may be missing, or other words (like “given (that)”) may stand in for the “if”. In all of these cases, the boundary between clauses that are joined by a conditional construction should be annotated with a clausal SU.

#### 4.5.3.2.1.4 Subordination in questions

Subordination can also occur within questions, as in the following example:

```
And if there's no inspectors /, what is the recourse /?
```

As with statement SUs involving subordination, these cases also invoke a clausal SU within the sentence.

#### 4.5.3.2.1.5 Embedded or multiple subordinations

While the examples above are relatively straightforward cases of subordination, more complex cases can lead to confusion. One such difficult case is that of additional subordination within the sentence. This is the cleanest of the difficult cases because we simply apply the rule for subordination anew and continue. In other words, if two or more clauses within a sentence stand in subordination, and one of those clauses (either main or dependent) itself contains another subordination, then this internal subordination is also signaled with a clausal SU. For instance, consider the following:

```
But more of them now seem to be coming at least to some middle
ground on all of this /, because they know the perils for them
politically /, if American young men and women are in harm's way
/& and they are then critical of the president /,
```

The application of this rule can continue ad infinitum, as long as the embedded constructions are themselves subordinations. For instance:

```
World Cup Football Champion France has said /, that it will not
take part in next year's Confederation cup in Mexico /, unless
the tournament date is changed /.

Today we took some important steps to ensure /, that Mr.
Milosevic knows /, that his indiscriminate use of force is
unacceptable /.

The administration had been on the record saying /, when all
```

```
of this came up in early November /, that if he didn't cooperate
one more time /, they would pull the trigger /.

Now the Falcons coach says /, he will modify his behavior /,
unless the game gets tight /.

They know /, if it goes to a vote in the Senate /, they can't
count on not having sixty-seven votes against the president /.
```

#### 4.5.3.3 Interactions between coordination and subordination

In many cases, subordination and coordination interact within a sentence, thereby complicating rules for SU annotation.

##### 4.5.3.3.1 Coordination in the dependent clause

In many cases of subordination, there not only exists subordination between the independent and dependent clauses, but also coordination within the dependent clause.

For instance, in this sentence:

```
If you write papers /& or you do design studies /, then it's real
hard to rate something like that /.
```

the dependent clause contains the coordination:

```
you write papers or you do design studies.
```

Under normal rules of coordination, we would expect to apply a sentence-level SU break between the two coordinated clauses, since they contain both subject and predicate. However, in this case, these clauses are part of the larger dependent clause caused by the if-then subordination. Rather than applying a sentence-level SU break, the two coordinated dependent clauses are separated by a coordinating SU break.

##### 4.5.3.3.2 Coordination in the independent clause

Likewise, in many cases of subordination the independent clause is itself a coordinated construction. For example, in:

```
The settlement requires /, that the company clearly state in its
mailings /, that no purchase is necessary to enter the
sweepstakes /& and it cannot tell consumers that placing an order
will improve their chances of winning /.
```

the independent clause is the coordination:

```
The company clearly state in its mailings that no purchase is
necessary to enter the sweepstakes and it cannot tell consumers
that placing an order will improve their chances of winning.
```

Similarly, the normal rules for coordination would impose a sentence-level SU break between the two clauses. Once again, the rule for coordination is overridden and we apply a sentence-internal coordinating SU. The result is that the entire subordination (the dependent clause and both of the conjoined clauses in the independent clause) form a single sentence with one sentence-final sentence-level SU break.

Another example follows:

```
If they've given up on the inspection regime /, then we're
talking about either containment and perpetuity which I think is
unreasonable /& or we're talking about some sort of action which
will lead to the removal of Saddam Hussein /.
```

### 4.5.4 Additional SU considerations

#### 4.5.4.1 Conversational introductions

It is common in the MDE telephone data for conversations to begin with an utterance that establishes common ground between the two speakers with respect to the assigned topic of conversation, e.g.,

```
So, air quality.
```

Such utterances should be classified as a statement SU.

#### 4.5.4.2 Progressive and participial phrases

Sometimes speakers use a clause beginning with the progressive or participial form of a verb as a kind of modifier. (This happens most often in Broadcast News but can also happen in telephone speech files.) In these cases (unlike the somewhat similar relative clauses, Section 4.5.4.6 below), a clausal SU break is applied before the progressive form of the verb. For example:

```
In Washington it'll be back to work for the president /,
preparing for the State of the Union address later this month /.

He's up there moving around /, smiling like he's happy /.

Cousteau set off in April /, accompanied by his longtime crew /.
```

#### 4.5.4.3 Prefacing statements

Speakers often preface opinions and other statements with a kind of term, usually ending in a “to be” verb (optionally followed by “that”). These are difficult to characterize precisely, yet they occur with some frequency in conversational speech. For instance:

```
The first thing for me is /, I see a couple of different ways of
talking about what privacy is /.

Another thing that occurred to me is /, there's not so much
invasion of my privacy /.
```

```
The question is /, did he really know it was going to be good /?
Or did he just do it /?

What we run into is /, we have the Texas Air Control Board TACB
that send out jurisdictions under which we have to reply to /.
```

These prefacing statements will receive a clausal SU break.

#### 4.5.4.4 Refinements and generalizers

Refinements follow an original statement with some additional or revised information. They should not be confused with edit disfluencies that contain a depod. Refinements are tagged with a clausal SU break, as in these examples:

```
Right now I believe there are two policies that the government
uses to modulate immigration /, three actually /.

We went camping this past weekend with some friends that had a
minivan /& and pulled the little trailer behind /, the pop up
trailer /.
```

Generalizers are phrases like “stuff like that” or “and so on” that are used by a speaker to extend their specific examples to a more general observation. They often follow lists. Generalizers invoke a clausal SU break, for instance:

```
Asleep At The Wheel, Sons of the Pioneers /, stuff like that /.

In the summer I like one piece dresses, short sleeves /, things
like that /.
```

Generalizers often occur alongside refinements, as in this example:

```
And he wanted to drive that van /& and travel /& and wanted
everything in there /, T V /, the whole bit /.
```

#### 4.5.4.5 Quotes

Quotations are a very frequent occurrence, particularly within broadcast news data. Whether direct or indirect, quotations impose a clausal SU break. For instance, in the following sentences:

```
But the FDA says /, the drug can have a dramatic impact on people
who suffer from Crohn's disease /.

Buccaneers coach Tony Dungy says /, that even if Tampa Bay does
not make the post season /, the team is playing the kind of
football it likes /.
```

the indirect quote causes a clausal SU break after the quotative verb (say, say that). The clauses remain joined in a single sentence with one sentence-level SU boundary.

Direct quotes follow the same basic guidelines as indirect quotes. A direct quotation imposes a clausal SU break after the verb of quotation:

> When it was over, Samuel Haile, a fifth-grader, said **/**, "I want to learn more about John Glenn." **/.**

When the attribution follows, rather than precedes, the initial quote, the clausal SU break occurs directly after the quote:

> "There was a President of the U.S. who really made a commitment to the country that before the end of the 20th century, we were going to have a man on the moon," **/**, Principal Plaut said **/.**

(Note that the sentence-level SU type should be assigned based on the sentence as a whole rather than just the quoted material, so that a sentence like the following

> The principal then asked **/**, "Does anybody know who that president was?" **/.**

receives an sentence SU label of statement rather than question.)

When the quoted material continues for several sentences, additional SUs may be necessary, as in the following example:

> "Gerry has got that revolutionary aura, that whiff of cordite," **/**, says one friend **/.** "Bianca Jagger calls him **/.** All the rich American society hostesses throw themselves at him **/.** He's amused by it." **/.**

All of the standard rules for annotation of complex SUs apply within quotations, whether indirect or direct.

#### 4.5.4.6 Relative Clauses and *where* modifiers

Relative clauses modify a Noun Phrase (NP), providing more information about the NP. They are often signaled by the words “who”, “which” or “that”. Relative clauses **do not** motivate an SU boundary.

> He usually goes to the store ∅ that has the good ice cream **/.**
>
> I now have a letter from Trent Lott and other Senate leaders, including Jesse Helms, ∅ who is the chairman of the Senate Foreign Relations Committee **/.**
>
> Passengers ∅ leaving on Flight738 reported strong turbulence **/**, before the plane went down **/.**
>
> And we are presently in receipt of a site permit ∅ which will allow us to have certain emissions up to a certain tonnage **/.**

Similarly, "where" can begin a phrase that modifies a previous NP. These "where" clauses should be treated like relative clauses, receiving no SU break:

```
But in reality I think /, what you would wind up with is a
political football Ø where they would see all these body counts Ø
that they can use for their own will /.

We both were raised that way Ø where our parents sort of stuck us
in front of a TV to be the babysitter /.
```

#### 4.5.4.7 Broadcast news conventions

Stylistic conventions in the reporting of broadcast news lead to challenges for SU labeling. Broadcasters often employ strings of sub-sentence units rather than longer complex sentences. For the most part, these shorter units will be treated as separate SUs. Keep in mind that punctuation in these cases is highly variable and should not be relied upon to provide meaningful insight into the proper identification of SUs.

##### 4.5.4.7.1 Introductory headlines

These often take the form of a series of brief phrases:

```
Now that El Nino is virtually gone there is La Nina to worry
about /. One hot and one cold /. We'll take a closer look /.
```

Sometimes a statement (like a story title) is followed by a question:

```
Airline passengers and outrageous behavior at 30000 feet -- /.
what can an airline do /?
```

These brief bullet-point phrases or sentences should generally be treated as individual SUs because they can stand alone and express a complete idea.

##### 4.5.4.7.2 Formulaic constructions

Broadcast news reports often include short "bylines", or formulaic introductory announcements consisting of e.g., the date, program title, announcer's name, listing of today's top headlines and the like. Typically, each of these short phrases constitutes a separate SU:

```
Richard Harris /. NPR News /. Washington /.

And this is VOA News Now /. I'm Neil Curry with Theresa Erikson at
22:53 Universal Time /.
```

When the reporter signoff consists of a well-formed sentence (a complete main clause) it should be tagged as a single (statement) SU:

```
I'm Angela Astore reporting from Dallas for CNN /.
```

A similar construction is used when one broadcast reporter turns the floor over to someone else by saying the person's name. Anchors and reporters often finish their story introductions and conclusions by adopting this technique, e.g.,

```
Finally, the governor's stance on rising taxes /. We go to Angela
Astore in Dallas /. Angela /.

The governor said he would not comment on tax cuts /. Peter /.
```

These cases merit a statement SU break, even if the speaker's intonation rises on the name. Since this construction is merely a way of passing the turn to the next person and is not an actual question, the annotator should not tag the name with a question SU break.

On the other hand, a reporter will often *preface* the initial statement of a report with the anchor's name. In the case of prefacing, the name does **not** merit a separate statement SU break, and is considered a part of the larger complete SU:

```
A: We go to Angela Astore in Dallas /. Angela /.
B: Peter ∅ the situation here is hectic /.
```

#### 4.5.4.7.3 Multiple overlapping speakers

Because broadcast data includes many speakers on a single channel, there are occasionally periods of overlapping speech that involve multiple overlapping speakers. In these cases, SU assignment can be quite difficult because it may be unclear when there is an interruption versus two speakers simply talking over one another. Annotators should use the "NoRT Annotation" label (fully described in Section 9 below) whenever a span of speech/transcript is too difficult to annotate accurately.

### 4.5.5 SUs across speaker turns

Complete SUs can continue across speaker turns. This happens most frequently in the case of failed interruptions, as in this example:

```
A: we did get the wedge cut out by building some kind of
B: a cradle for it /.
A: cradle for it /.
```

## 4.6 Deciding among SU types

In some cases it may be difficult to decide among SU types, depending on the function of the SU within the discourse. Annotators should adhere to the following hierarchy:

| **Sentence-external** | **>** | **sentence internal** |
|---|---|---|
| Incomplete SU > | | Coordination > |
| Statement > | | Clausal |
| Question > | | |
| Backchannel | | |

Sentence-external SU breaks are the default; that is, if there is motivation for a sentence-external SU break it should be assigned over and above a sentence-internal break. For instance, in the following utterance a question SU "wins out" over a coordination SU in the first clause:

```
Who is he /? and what is he doing /?
```

Within external and internal SU breaks, different break types are also ordered. Given this hierarchy, a backchannel that is spoken with a question intonation (rising final intonation) and functions as an interrogative should be labeled as a question, not a backchannel.

```
A: I almost got hit by a bus once on my way to work /.
B: Yeah /?
```

## 4.7 Interactions between SUs and other types of metadata

### 4.7.1 Fillers and SUs

Fillers often occur at the edges of SUs. In the case of a "trailing" filler, the filler comes at the end of turn that includes a full statement, question or other sentence-level SU. For these "trailing fillers", the filler should be excluded from the span of the sentence-level SU. When the same speaker continues in the next turn, the filler will be captured in the span of the following SU. Otherwise, the "trailing" filler is treated as an incomplete SU. For instance,

```
A: I'm usually really insulted by them /, so
A: I just hang up as soon as I recognize what they are /.

A: Yeah, that was a lot to go through /. so /-
B: You must be relieved to be over that /.
```

Fillers (including asides/parentheticals) that come at the beginning of larger sentence-level units are captured within the following SU; no separate SU should be introduced after the filler.

```
See, ∅ my company has a much stricter policy than yours, it
sounds like /.
```

Similarly, when a filler directly precedes an edit disfluency but is not part of that edit, it is contained within the same SU as the edit (it should **not**, however, be labeled as part of the depod):

```
Um, ∅ [the, th-] the one thing I'm thinking is that it might be
hard to see the stage from way back there /.
```

Annotators should **not** look inside fillers for additional SU breaks. This is particularly relevant for lengthy asides/parentheticals. Like all fillers, these

should be wholly contained within the larger SU and should not contain any internal SU breaks:

> I read a thing **{I don't even remember if it was in the Dallas Site or the Inside one}** about companies allowing you to purchase extra vacation days**/.**

Occasionally, a filler will stand in isolation as a complete turn. In these cases only, the filler should be tagged as a separate statement-type SU:

> A: But it's just really bizarre, if you ask me **/.**
> B: **Um /.**
> A: The whole criminal justice system **/.**
> B: **Um,** ∅ but I don't think the police are the biggest problem **/.**

In such cases annotators must be careful to distinguish fillers from backchannels.

### 4.7.2 Disfluencies and SUs

Edit disfluencies are always contained within a larger SU rather than standing alone as a single unit:

> [Her] * all her kids were sick **/.**

Furthermore, annotators should not indicate SU breaks of any kind within the depods they identify. (See Section 4.4.4.1 for a related discussion of restarts and Incomplete SUs.)

## 4.8 SU conclusions

Most SUs will be easy to identify and classify, but there will be numerous cases of ambiguity, particularly in spontaneous speech. Annotators should keep in mind the ultimate goal of metadata annotation: to enhance transcript readability. When in doubt about where to place an SU boundary, annotators should rely primarily on the semantic information conveyed by the utterance and should apply the SU break in accordance with the primarily syntactic rules detailed above. Prosodic features, while interacting with SU boundaries, play only a limited role in SU assignment.

# 5 Deciding among MDE types

## 5.1 MDE annotation type hierarchy

Occasionally, annotators may have difficulty deciding which metadata type to assign to a particular utterance or token. In such cases, the following hierarchy holds:

DM/FP/EET > SUs > A/Ps

In ambiguous cases, annotators should use this hierarchy to decide how to tag an utterance.

### 5.1.1 Asides/Parentheticals vs. complete SUs

One particularly difficult decision point results from the current treatment of asides and parentheticals (see Section 2.4 for discussion). Although A/Ps are usually semantically "off-topic", decisions about whether to treat an utterance as an aside should be informed by its syntactic status as well. Asides interrupt a larger SU and are nested within it. A phrase that can stand alone as a statement SU and does not interrupt a sentence-level SU should be tagged as a separate SU, not as an aside. For instance, the following are examples of an aside:

```
He has now for about {oh gosh how long has it been} ten years I
guess

It's really {I don't know} incredible the way it all happened /.
```

while these are not:

```
He has now for a while /. Oh gosh how long has it been /? Ten
years I guess /.

It just all seems so crazy /. I don't know /. I guess it's going
to be fine /.
```

# 6 Noises

Speaker noises like **breath**, **cough** and **laugh** as well as background **noise** tags appear with some frequency in the transcripts. These tags appear in **light green** font in the annotation tool. Given a more robust annotation tool, noises would be masked from the transcript file because their presence has no bearing at all on the metadata task as currently defined. As this is not currently possible, as a rule of thumb annotators treat noises like fillers, excluding them at the edges of depods and SUs. However, annotators should not spend any time considering how to annotate noises "correctly" but should do whatever comes naturally.

# 7 Transcription errors

As annotators label metadata in the existing speech and text files, they are likely to encounter transcription errors in the original transcript files. Annotators are not permitted to edit or repair existing transcripts[5], but should label the extent of a questionable transcription as such so that these errors can be readily located and corrected later.

It is likely that many of these questionable transcriptions will occur in areas of disfluency, since these are notoriously difficult to transcribe. **In such cases, the annotator should add metadata annotation, taking the transcript file (not the audio) as accurate.** The entire extent of the region of questionable transcription should also be identified as such using the “questionable transcription” tag.

[5] Transcript correction is not possible under the current task definition. This issue may be revisited in future MDE efforts.

# 8 "Difficult Decision" label

Some annotation decisions will be harder than others, and annotators might end up spending an inordinate amount of time deciding how to handle a small number of very difficult cases. In order to avoid spending too long on any one annotation decision, annotators should use the "difficult decision" label to indicate that a given decision was very hard to make. This label can apply to any type of annotation decision. Because difficult decisions are also likely to have lower annotator confidence, all "difficult decision" annotations will be revisited during a separate quality control pass over the data.

# 9 "NoRT Annotation" label

While the "difficult decision" label is used when the annotator is unsure about whether the given annotation is correct, in some cases it may be impossible to make any annotation decision at all. When the audio signal is distorted, when there is a great deal of overlapping speech that makes it impossible to sort out who is saying what or where interruptions vs. incomplete SUs occur, when the transcript presents a commercial or other material that should not be annotated, or when the transcript is so inaccurate as to inhibit understanding, the annotator should select the entire span of the problematic region and label it with the tag "NoRT Annotation". This label should also be used at the beginning or end of a file when a particular turn or SU is artificially truncated. Regions labeled " NoRT Annotation " should receive no annotation whatsoever.