Dysfluency Annotation Stylebook for the Switchboard Corpus

Marie Meteer, et al.
revised by Ann Taylor

Contents

1 Introduction

The Switchboard Corpus comprises telephone conversations between two individuals regarding a specific topic. Since it is conversational speech, it contains fragments of words, interruptions, incomplete sentences, fillers and discourse markers which require annotation according to specific and consistent rules.

Prior to discussing the different kinds of annotations required, it is important to briefly describe the corpus. As mentioned earlier, the corpus essentially comprises conversations between two speakers, specified as speaker A (A:) and speaker B (B:). A sample subset of a conversation is as given below.

```
A:  You interested in woodworking?

B:  Yeah.  Actually, I, uh, I guess I am <laughter>.  Um, it just seems kind
of funny that this is a topic of discussion.  Uh, I do, uh, some, uh,
woodworking myself <noise>.  Uh, in fact, I'm in the middle of a project right
now making a bed for my son.

A:  Um.  #What kind of# --

B:  #It's, uh,#

A:  -- pine?
```

As can be observed from the above conversation, there are other markings besides textual English. Words in angle brackets (<...>) indicate non-speech events, ranging from laughter, noise, breathing, lipsmack to TV and baby_crying. All these words are considered as a part of the conversation although they are not used for annotation purposes. Besides these non-speech events there are other notations frequently used by the transcribers, such as <<faint>> or <<very faint>> (which implies that the word could not be heard clearly), <<mispronounced>> etc.. These marking have been left in the text to help the annotator in interpreting the transcription of the conversation.

Fragmented or incomplete words are marked in the transcription with -, for example, *Ut-*, or *you kn-*. A few other markings that are present in the transcriptions include −− at the beginning of a turn, indicating a continuing sentence across the turn and #...# which indicates two speakers overlapping each other.

When sentences are separated, the restarts folded and the non-sentence elements removed, what remains should be reasonably grammatical sentences. Though the "cleaned up" sentence should be reasonably grammatical, they need not be complete sentences, particularly in the case of replies, acknowledgments or sequences of words that serve to continue the conversation, as long as they are **complete constituents** standing alone (see Section 3.2 for details).

The example given below illustrates a subset of a conversation, which is annotated and later cleaned based on the annotations. All three versions of the example, namely the original, annotated and cleaned versions, are given below. Note that the *cleaning* of the annotated example is just an indication of what *could* be done and not necessarily what *will* be done.

Original:

A: he's pretty good. He stays out of the street and, uh, if I catch him I call him and he comes back. So he, he's pretty good about taking to commands and --

B: Um.

A: -- and things.

B: Did you bring him to a doggy obedience school or --

A: No --

B: -- just --

A: -- we never did.

B: -- train him on your own and,

A: I, I trained him on my own and, uh, this is the first dog I've had all my own as an adult.

B: Uh-huh.

Annotated:

A: he's pretty good. / He stays out of the street / {C and, } {F uh, } if I catch him I call him / {C and } he comes back. / {D So } [he, + he's] pretty good about taking to commands [and + --

B: {F Um. } /

A: -- and] things. /

B: Did you bring him to a doggy obedience school or --

A: No -- /

B: -- just --

A: -- we never did. /

B: -- train him on your own / {C and, } -/

A: [I, + I] trained him on my own / {C and, } {F uh, } this is the first dog I've had all my own as an adult. /

B: Uh-huh.   /

Cleaned:

```
A: he's pretty good. He stays out of the street. if I catch him I call him. he
comes back. he's pretty good about taking to commands and things.

B: Did you bring him to a doggy obedience school or just train him on your own

A:  No we never did. I trained him on my own. this is the first dog
I've had all my own as an adult.

B:  Uh-huh.
```

An important point to be noted while annotating the conversations is that the two speakers stand independent of each other; namely only Speaker A can complete his/her sentences though Speaker B might appear to complete them. This point will be clearer as we discuss the sentence annotations in Section 3.2.

2 A Quick Reference Section

The following chart gives an overview of the annotations.

Slash-units:	complete	: boundary marked with /
	incomplete	: boundary marked with -/
Restarts:	restarts with repair	: [RM + RR]
	restarts with repair and non-sentence elements	: [RM + { } RR]
	restarts w/o repair	: [RM +]
	complex restarts	: usually go from left to right.
Non-sentence elements:	filler	: {F...} (*uh, um, huh, oh,* etc.)
	explicit editing term	: {E...} (*I mean, sorry, excuse me,* etc.)
	discourse marker	: {D...} (*you know, well, so, like,* etc.)
	coordinating conjunction	: {C...} (*and, and then, but, because,* etc.)
	aside	: {A...} (aside within sentences/restarts)

3 The Annotation System

This section is divided into three parts:

- Non-sentence elements (3.1)
- Slash-units (Section 3.2)
- Restarts (Section 3.3)

3.1 Non-sentence elements

Non-sentence elements are dysfluencies which are divided into a number of types and labelled according to the following classification scheme.

3.1.1 Filled pause {F...}

Filled pauses have unrestricted distribution and no semantic content. A few common examples are *uh, um, huh.* There are also rarer filled pauses, such as *eh, oops,* etc.. Each filler is enclosed separately in its own brackets. Note that *uh-huh* (whether spelled as one word or two) is not a filler, but *huh* is. (*Note: on second thoughts* huh *probably shouldn't be counted as a filler but rather an assessor/continuer*)

```
B: {D Actually, } [ I, + {F uh, } I ] guess I am <laughter>. / {F um, } it just
seems kind of funny that this is a topic of discussion. / {F uh, } I do, {F
uh, } some, {F uh, } woodworking myself <noise>. / {F uh, } in fact, I'm in
the middle of a project right now making a bed for my son. /

A: {F uh, } {F uh } I see stuff in craft galleries for five hundred dollars /
```

When fillers occur at a unit boundary, follow the transcriber's punctuation in deciding which side of the boundary to put them. When a filler is separated from the end of a unit by a comma, consider the filler part of the previous unit. If it is separated by a period, however, consider it part of the following unit unless it ends a turn which does not continue in the next turn.

```
A.5: {F Uh, } no, / we're not being tested for drugs at all, {F uh, } / our
policies and procedures manual, {F uh, } the furthest it goes about drugs is in
[ the, + kind of the ] miscellaneous section, or --

B.76: {F Oh, } {C so } they don't go to school. / Is it like, {F uh, } -/ {F
oh, } what's that called, / it's, {F uh, } correspondence school, / they do it
--

B.30: Yeah, / [ it's really, + it's really ] bizarre. / {F Uh, } particularly,
{D like } where we are, {D you know. } /

B.12: -- I guess ] we can start. {F Uh, } /
A.13: Okay. /
@B.14: No, / [ I don't, + I don't ] have any kids. /
```

Oh is treated as a filled pause if it appears along with other words, for example *oh yeah, oh really.* When it appears as a response to a statement, (normally indicated by a following period), it is not counted as a filler, even if a new sentence follows in the same turn.

```
B:  {F Oh, } yeah. /  Uh-huh. /

A:  {F Oh, } is that right?  /

A:  {F Oh, } {F oh, } great.  /

A:  -- I'm in Dallas.  /
B:  Oh. /

A:  {D Well, } I saw [ one of, + some of  ] those early space capsules at
Smithsonian. / I'm sure you've --
B:  Oh. /
A:  -- seen more of them than I have,  /

A.102: I work [ at a, + for a ] university. {F Uh, } Georgia Tech, {F uh, }
research -- /
B.103:  Oh. / {D Well } good. /

B.34:  [ I'm, + I'm ] an electrical engineer. /
A.35:  Oh. /  In New Hampshire. /  # All right. # /

B.46:  Nobody came near it when he was there. /
A.47: Oh. / {F Uh, } [ I had, + I had ] a dog one time, / {C and, } {F
uh, } he chased [ after a, + after a ] skunk and got sprayed -- /
```

3.1.2   Explicit editing term {E…}

Explicit editing terms have some semantic content (e.g. *I mean, sorry, excuse me*) and usually occur between the restart and the repair. The exception is *I mean* which in addition to appearing in restarts also has a much wider distribution. (*Note: it isn't entirely clear that the* I mean *in restarts is the same as the* I mean *with wider distribution, but we have treated them the same*)

```
A: {F Oh, } yeah, / {F uh, } the whole thing was small and, [ you, + {E I mean,
} you ] actually put it on <laughter>, /

A: [ We've had kid, + {E or } we've had dogs ] when I was a kid, / {C but } [
this, + this ] is the first one that I, {F uh, } took in, /

B:  Yeah, [[[ I've, +  I've, ] +
A:  {D Well, } I, -/ {E excuse me, } go ahead.  /
B: {E I would say } I've never, ] + I've never ] seen an actual capsule. /
```

It is, however, possible for editing terms to occur outside the RR. This is most common for *I mean* but can occur with other editing terms.

```
B: {C And } sometimes, {D you know, } I don't even know how much of a
difference it really makes, {D you know. } /
A:  Yeah.  /
B:  {E I mean, }  I guess in some cases, {D sure } we care about what they
did in the past -- /

B:  -- {F uh, } how many people are self serving.  /
A:  Uh-huh.  /
B:  {E I mean, } you -- -/
A:  Uh-huh <<faint>>. /

B: -- I think it gets in the way of the better judgment. / {E I mean } just
because you did something in the past doesn't mean he's not a good enough
person to be a politician. /

A: Especially since my youngest one is, {E excuse me, } only fourteen months. /
```

I mean should always start a new unit, when it occurs at a boundary.

```
B.87: <phone_ringing> No matter what it is, / {E I mean, } [ I cou-, + I'm ]
perfectly capable of doing it for them, / {C but, }

B.163: Yeah, / [ plus, + plus ] I think it becomes [ part of y-, + part of you
] too, / {E I mean, } you get to stay home, / {E I mean } it's not just for
him, /
```

3.1.3   Discourse marker {D ...}

Discourse markers have a wider distribution than explicit editing phrases but have more semantic content than fillers (although not always much). Various discourse markers will be discussed below.

1. The two most common and easily recognized discourse markers are *well* and *you know.* When *well* occurs at a boundary, it should always start a new unit. The rules for positioning *you know* are given below.

```
B: {D Well, } we have a cat who's also about four years old.  /

A: {D Well } [ I'm, + I am  ] interested in woodworking. /  I don't have
lots of tools. / Our local schools have adult education in the evening -- /

A.47: Yeah, / {D well, } real interesting. /

B.88: Yeah, / {D well, } don't even talk to me, /

B: Yeah, / with, {D you know, } me being at home and just having the one
income, {D you know, } you don't have, this lot of extra money [ to, + to ] do
a lot of, {D you know, } extra things. /
```

```
A:  [ Maybe, + maybe  ] that's what is happening. / Maybe so many
things have become public with the media over the past, {D you know, }
ten to twenty years --
B:  Uh-huh.   /
A: -- {F uh, } parallel to the legal codes that, {F uh, } we're focusing on the
wrong thing. /

B: {C And } sometimes, {D you know, } I don't even know how much of a
difference it really makes, {D you know. } /
```

The discourse marker *you know* when it occurs at the boundary of slash-units should be put at the end of the first slash-unit unless the slash-unit ends in a period. However, when *you know* stands on its own as a turn, count it as the end of the previous turn, even if this ends with a period.

```
A:  -- you could see the scorch on them, {D you know, } / they had
though piles -- /

B: It's, ] + {F uh, } it's, ] {F uh, } plywood, {F uh, } face, I guess.  /
{C But } I cut it out in the shape of a crayon box / {C and then } I painted,
{F uh, } about eight different, {F uh, } colors, {F uh, } {D you know, } /
the crayons that are sticking up, it will be the headboard -- /

A:  -- it didn't, -/ {C but } if it cost me twenty dollars, {D you know, } /
# {C and then } I,# -/

B: {F uh, } I appreciate having the satellites.  / {D you know, } that's
certainly made it nice for other purpose. / I don't know if you have a
satellite dish or not. / We don't -- /

A:  {F Uh, } I'd say it's about an eighth.  / # {D You know } # jus-, -/

A: -- that, {F uh, } <breathing> how much, {F uh, } should it be our, {F
uh, } prerogative to see into the private lives of these people.
B:  Yeah.   /
A:  {D you know. } /
```

2. *Like* can be a preposition or a discourse marker. Unfortunately distinguishing the two can be difficult without intonation cues, leaving the possibility of ambiguous cases in the transcripts. Compare *They're like bermuda shorts* and *They're, like, bermuda shorts.* In the first example the shorts are NOT bermudas, only like them, while in the second the shorts ARE bermudas. Context and the transcriber's use of commas may help in these cases, but some cases will probably remain ambiguous. In this case, if the speaker is a heavy discourse *like* user, count ambiguous cases as discourse markers, if not, assume they are not.

```
B: they thought that was lot of fun. / You can have
birthday parties there / {C an, } I think that was better than {D like }
Showbiz Pizza cause there's more for them to do. /
```

3. *So* is another variable case. For *and so* see Section 3.1.4. (*Note: annotators were basically unable to distinguish the discourse marker from the conjunctive use of* so)

There are three sentential uses of *so* in the texts:

(a) as a subordinating conjunction where it is equivalent to *so that* with purpose meaning.

```
I went and bought groceries so there'd be something for dinner.
```

(b) as an adverb meaning *therefore* and/or more or less replaceable with *and*.

```
I wanted to get into a good college, so I studied real hard in school
```

```
I grew up on a farm so I always had outdoor pets
```

(c) as a sort of sentence introducer (basically any initial use that doesn't fit into either of the other two categories, also often (but not necessarily) followed by a pause (indicated by a comma).

```
So, do you like it?  So, this is how we do it.
```

The first use is not marked in any way and the clauses introduced by it are not separated from the main clause, if the main clause is present. The second example gives a case where *so* (in B's second turn) expresses purpose and so falls into the first category, but the main clause that it is subordinate to is expressed by the other speaker and thus the *so*-clause stands on its own.

```
B: ... {C and, } {D you know }, teach them about their grandparents or
something like that. /  Bring out pictures of them so they get to know them /
{C and }, we just do all kinds of fun things like that /

A:  {D So } we did get the wedge cut out by building some kind of --
B:  A cradle for it. /
A:  -- a cradle for it. /
B:  Yeah, / so you can steady it  / {C and then, } you used a handsaw
or a backsaw? /
```

The second use is marked as a coordinating conjunction and begins a new slash-unit.

```
B: Yeah, /  yeah, / yeah, / I had, {F uh, } a similar, {F uh, } experience. /
I  grew up on a farm / {C so } I always had, {F uh, } outdoor pets /

B: {D Well, } {F uh, } [ it's, + we ] just moved recently <laughter> / {C so}
now we're in the, {F uh, } Dallas area / {C and } it's [ very + --
A:  Oh.  /
B: -- very ] nice / {C and, } -/

A: [ We've had kid, {E or } + we've had dogs ] when I was a kid, / {C but } [
this, + this ] is the first one that I, {F uh, } took in, / {C so } I wasn't
sure if I'd be able to get it all right the first time, / {C but, } {F uh, } [
he, + he ] seems to have picked it up pretty well. /

A: Yeah, / {D well } , a friend of mine at work here said that he tried it with
his dog / {C and } he wanted to see what he was subjecting his dog to / {C so }
he held on to the collar / {C and } he walked out to the perimeter.  / He said
it was a good jolt. /
```

The third use is labelled as a discourse marker.

```
A:  {D So, } I'm a licensed teacher / {C but } jobs being what they are, I
    substitute  teach -- /

B:  {D Well } <breathing>, {F uh, } <breathing> <noise> I'm not sure to be
    honest with you, / {D so, } {F uh, } this going to the moon seemed to be
    a exciting adventure / [ {C and, } + {F uh, }
A:  Ego trip.  /
B: {C And, } ] yeah,/ it was nice to go there, / {C but } I don't know that
    it really benefitted mankind that much. /

B: {D So } yeah. / to do [[ the, + {F uh, } th-, ] + the ] turn
    cut. / {F uh, } how about yourself? /
```

As mentioned above the annotators were not able to distinguish the types of *so* very well. The
following rules were adopted part way through the project in an effort to make the bracketing of
so more consistent.

When *so* occurs trailing the end of a unit and there is no continuation in the following turn, it is
marked as a conjunction.

```
B.30: -- {C and } [ it, + it ] didn't even hardly --
A.31: {F Oh, } didn't really even count. /
B.32: -- go through, / {C so. } -/
A.33: Yeah. / [ I'm, + I'm ] a college student / {C so } I haven't been, {D you
    know, } a jury eligible age for very long -- /
B.34:  Yeah. /

A.55: [ In, + in ] the case you were involved in, you said it was just sort of
    a bank matter of some kind. / {C So. } -/
B.56: Yeah, / [ the, + the ] bank was suing them [ for, + {F uh, } ] because
    they went to get [ the, + the ] trailer --
A.57:  Yeah. /
```

When *so* occurs alone as a turn with no continuation in the next turn it is bracketed as a discourse
marker. In retrospect, it might have been better to do these as conjunctions as well, but since
there's no context in these cases the distinction is impossible to make.

```
A.81:  I think for a lot of people it has. /
B.82:  Certainly. /
A.83:  {D So, } -/
B.84:  Yeah. /
A.85:  {D So, } -/
B.86: {C But, } {F uh, } {F uh, } I know that, {F uh, } in some of these
    larger cities they're passing out free syringes and all of that sort of
    thing --
A.87:  Uh-huh.  /
```

Otherwise, the rule is to label *so* as a conjunction when it prefaces a sentence which continues
the previous narrative and as a discourse marker when it introduces a change in topic or a radical
change in the previous topic.

Note that while *and so* is bracketed as a compound conjunction, *and, uh, so* was bracketed as a
restart. On the other hand, when *and* and *then* are separated by a filler, only *and* is bracketed as
a conjunction and *then* is left bare.

```
{C and so }

[ {C and, } + {F uh, } {C so } ]

{C and then}

{C and, } {F uh, } then
```

4. *Actually* is also sometimes a discourse marker but sometimes just acts as an adverb (much like *well*) although it is more difficult to distinguish the two in the case of *actually*. Discourse *actually* tends to be initial or in an initial string of non-sentential elements, and is often followed by a pause, indicated by a comma. (*Note:* actually *also proved impossible for the annotators to mark consistently and was jettisoned as a discourse marker part of the way through*)

```
A: [ He, + {D actually, } he ] just went to the vet yesterday to get his shots
and everything, / {C and } they put him on the scale / {C and } he's up to
ninety four now. /

A:  -- usually. ] / [ I'm, +
B: [ What, + what  ] type of art do you focus on?  /
A:  {F uh, } {D well } {D actually, } I'm  ] very, {D you know, } -/ I have a
Master's degree / [ {C and, } +  {C and } ] my projects had to do with
acrylic painting, / {C but } I'm really interested in the crafts. /

A:  You interested in woodworking? /
B: Yeah.  / {D Actually, } [ I, + {F uh, } I ] guess I am /. {F um, } it just
seems kind of funny that this is a topic of discussion.  / {F uh, } I do, {F
uh, } some, {F uh, } woodworking myself /.  {F uh, } in fact, I'm in the
middle of a project right now making a bed for my son. /
```

5. Other items, usually adverbs, are also sometimes used as discourse markers. These cases can usually be distinguished because the lexical item in question is utterance initial and does not have its usual meaning and/or is rather bleached of meaning (sorry this is a bit vague). The two most common are *now* and *see*.

```
A.23: {D Now } we've let some people go, {F uh, } that [ wa-, + we, ] [ was
+ ] found that drugs were used in their system. / {D Now, } I don't know
how T E C would handle it. / They haven't pursued it, {D you know. } /

B.34: {C But, } {F um, } no, / I, {F uh, } -/ my step classes and toning is
about the only thing / {D well, } {E I mean } ... -/
A.35: {D Now } [ what's, + what's ] toning? / Is that [ lift-, + lifting ]
weights? /

B.30:  {F Um, } {D well, } {D see, } I'm a school principal. /

B.36: -- in spite of everything we can do / {C and, } {D see, } I'm afraid
I think if we would take the equivalent amount of money and invest it in
young people, that, -/ {E I mean, } course, you couldn't do that because
you got to do something with the ones that are already there -- /
```

12

Discourse markers can occur within a restart, after the IP, as with other dysfluencies.

```
B: [ We take, + {D you know, } whenever we take ] them to Showbiz or -/ they
think it's wonderful just to go to McDonalds, /

B.64: If you compare [ the, + {D like } the ] people [ that could have, + that
could have ] sent their kids to private schools and [ the people that, + {D you
know, } the people that, ] {F uh, } did send their kids # to private # --
```

3.1.4   Coordinating conjunction {C...}

Coordinating conjunctions include at least the following: *and, and then, and so, but, but anyway, either, or, because* (and variants *cause, coz,* etc.), and *so. So* and *because* are only counted as coordinating conjunctions in certain of their uses. The use of *so* is discussed in Section 3.1.3. *Because* is discussed below. (*Note:* but anyway *is more likely to be bracketed as* {*C but* } {*D anyway* } *or only the* but *may be bracketed*).

```
A: Yeah, / {C and } we got him when he was about eight weeks old / {C and }
he's pretty okay, /

A: {D well } {D actually, } I have a Master's degree / [ {C and, } +
{C and } ] my projects had to do with acrylic painting, / {C but } I'm
really interested in the crafts. /

A:  {D Well, } I thought there was some medical benefits that we have
learned from space, / {C but } did he go into that in the book? /

B: {C and then } I painted, {F uh, } about eight different, {F
uh, } colors, /

A:  {C But anyway, }  what do you think we've gained from the
spaceflights? /

A: {C And } he usually is good about staying within them, although our next
door neighbors have a dog, too / {C and, } {F uh, } [ she, + she ] is good
friends with my dog. /
B: {F Oh, } yeah?  /
A: {C And so } he often gets to smelling her scent and will go over there
to sniff around and stuff /

B:  {F Oh }, {D well } <laughter>, I just got through [ renting, +  reading,
 ] {F uh, } SPACE by, {F uh, } James Michener. /
A:  {F Oh, } is that right?  /
B:  {C And so } that, {F uh, } helped out a little bit I think.  /

A: Have you ever gotten one of those calls that is either generated by a
computer or somebody going down a list / {C and } their {C either } offering a
```

```
service / {C or } they're introducing some new product in the area / {C and }
normally when they call, you're {C either } in the shower, / {C or } you're in
the middle of cooking something / {C and } you have to stop everything to run
to the phone. /
```

In its normal causal subordinating conjunction role, *because* is not marked in any special way. There is, however, another use of *because* where it is not particularly associated with any main clause, but rather means something like *The reason I'm saying this is.* In this case it is annotated as a coordinating conjunction (CC). These cases often (but not always) start a new turn. (*Note: the annotators had difficulty distinguishing the two types of* because *and tended to overuse the conjunction type*)

```
A.  -- {C and } I went / {C and } I signed up for class because then again, you
[ get, + {D you know, } get ] someone to guide you / {C and } you get the tools
you don't --

A.  {D Well, } yeah, / [ I, + I'm ] signed up for a class in pottery making
because I've never tried that yet -- /

A: That's probably a good idea.  / They're good though.  / {F Uh, } have you
ever had a Retriever before? /
B:  No,  / {C but } my husband had one when he was growing up.  /
A: {C Because } [ I, + I ] find that they're good around other animals like
cats and other dogs.  / They don't tend to get real jealous or territorial.  /
They like to play. /

A: That's right <laughter>. /  {D Well, } do you think we should continue to
support it at the level, / {D well, } I thought probably one of the things they
were doing that they didn't tell us was spying.  / {C Because } --
B:  Uh-huh.   /
A: -- I think those satellites, {D you know, } or not satellites, but the
spaceflights could really spy /
```

Dangling CCs, if they are followed in the next turn by a clause, are treated as the beginning of a new slash-unit and bracketed. If they are followed by an NP in the next turn, they are not bracketed as CCs (see the first example). When a CC takes up a turn by itself and there is no continuation in the next turn, the CC is bracketed and the turn is labelled as an incomplete slash-unit.

```
A:  {C And so } he often gets to smelling her scent and will go over
there to sniff around and stuff / {C but, }
B:  <Laughter>.
A:  he's pretty good. / He stays out of the street / {C and, } {F uh, } if I
catch him I call him / {C and } he comes back. / {D So } [ he, + he's ]  pretty
good about taking to commands [ and + --
B:  {F um. } /
A:  -- and ] things.  /

A:  {C And, } {F uh } --
```

```
B:  Uh-huh. /
A:  -- {D you kn-, } when it becomes public, though, it becomes
scandalous. /

B:  Right. / I've seen them / {C and then } the watermelon, the red [ with
the, + with the ] seeds painted in  / {C and, }  -/

A:  -- {C but, } -/

A:  <Lipsmack> {C But, } -/
```

3.1.5   Aside {A...}

This is a category for "asides" that interrupt the flow of the sentence. Asides are rare and only occur when the corresponding sequence of words interrupts the fluent flow of the sentence AND the sentence later picks up from where it left off. Most examples of asides are found in repairs. Asides which are essentially comments on some part of the sentence are not bracketed as such.

```
B: I, {F uh, } talked about how a lot of the problems they have
to [ come, + overcome ] [ to, + {F uh, } {A it's a very complex, {F uh, }
situation } to ] go into space. /

A: {D So } we built a cradle for it / {C and } [ we got
th-, + {A once it was turned, } we got ] [ one s-, + one ] cutout [ on
the table saw, + on the radial saw, ] /

A:  {C And } I couldn't help thinking [ when that last quest-, + {A it was a
funny question  } --
B:  Uh-huh. /
A:  -- when that last part of the question  ] came about [ h-, + how  ]
many things can we, {F uh, } actually put a legal code on -- /

A:  -- I think those satellites, {D you know, } {A or not satellites, but the
spaceflights } could really spy /

B.46: ...  It's, I don't know, fifteen minutes from the Mass-, border and five
minutes from the main border. /

A.85: ... {C and } our neighbors were, I don't know, kind of slimy, /

B.8: ... {C And } in civil law there, {D you know, } is such a thing as {D like,
} let's say, {F uh, } <breathing> misappropriations or misuse of financial
instruments or something like that. /

A.11:  {F Uh, } <lipsmack> {C but } I do agree that, {F uh, } generally
[ it's, + society has ] sort of, {F uh, } let's see, rushed everything
ahead. /

B.16:    -- you think that the individual has as much time as they did,
```

```
let's say, ten, twenty years ago? /

B:   -- {F uh, } I know some. / Unfortunately, some elderly [ have to go to + {E
or } have, ] I guess it's Alzheimer's or # where # --

B.51: -- {C and } [ he, + he ] has now for about, oh gosh, how long has it
been, ten years I guess. /
```

The following are counted as asides although it isn't clear that this is the right analysis.

```
A.4: I don't know if it's even true that it's always unanimous because I
thought there were cases [ where, + {F uh, } {A I don't know if it's the
difference between felonies and misdemeanors, but } where ] it was okay for a
state to have it {D like } eleven out of twelve. /

B.16: ... {C but } what they do is employ, {F uh, } [ homeless, + {A I think
mostly men, but } homeless ] people to {F uh, } go out and collect all the
recyclable cans /

B.83: {F Huh. } / {D Now } I've done it [ with, + ] where [ [ you put, + you
cook the ri-, ] + {A well, I guess you probably could do it same way, but }
cook the rice ] and then take a can of {D like } cream of mushroom soup or
cream of chicken soup, and put a little bit of white wine in it. /
```

3.2   Slash-units

A slash-unit is maximally a sentence but can be a smaller unit. It is common in speech to proceed in a somewhat telegraphic manner so that many of the turns do not consist of complete sentences, but are nevertheless felt to be complete utterances. It also equally common for a speaker to stop an utterance in midstream, as for instance when interrupted. Intuitively, slash-units below the sentence level correspond to those parts of the narrative which are not sentential but which the annotator interprets as complete. Incomplete slash-units (see Section 3.2.2) are the segments of a turn where the speaker seems to have stopped in midstream and not completed an utterance she/he had started on. Any turn consisting of only continuers or assessments (expressions such as *uh-huh, right, yeah, oh really*) is also coded as a slash-unit. In a series of continuers/assessments, each one constitutes its own slash-unit. Interjections like *wow, boy, man, gosh, my goodness*, etc. do not constitute a slash-unit on their own unless they occupy a complete turn which is not continued in the next turn.

1. Continuers, etc.

    ```
    B:   Right. /

    B:   Uh-huh. /

    B:   Oh. /

    B:   Uh-huh, / yeah.   /
    ```

```
A:  Yeah, /  right.  /

B.55:  {C but, } boy there's a lots going on in there <laughter>. /

B.119:  # {F Oh } boy, # # that's great. # /

B.79:  ... {C and } he said {D well } gosh, you're just healing really well, /

A.74:  ... gosh I'm sure, not even one percent of the world is as fortunate as
these areas /
```

2. Phrases

```
B:  Yeah. / Kind of jack of all trades, master of none. /

B:  A band saw? /

B:  A cradle for it. /

A:  Especially with the kids. /

A:  The very first ones, / the little bell type things.  /
```

3. Simple sentences

```
A:  Uh-huh. /  I guess he's young. /

B:  He's about five and a half years old. /

A:  I went there, /  we have Home Depot out here. /
```

4. Subordinate Clauses standing alone

```
A:  -- it didn't, -/ {C but } if it cost me twenty dollars, {D you know, } /
# {C and then } I ,# -/

A:  {D So } we did get the wedge cut out by building some kind of --
B:  A cradle for it. /
A:  -- a cradle for it. /
B:  Yeah, / so you can steady it  / {C and then, } you used a handsaw
or a backsaw? /

B: {D So } yeah. / to do [[ the, + {F uh, } th-, ] + the ] turn
cut. / {F uh, } how about yourself? /

A: ... I don't think they're always necessary. / If you put enough patience
into, -/
B:  Yeah, / just be consistent and diligent --
A:  Uh-huh.  /
```

17

```
B:  --  with it  / {C and, } {F um, } -/
```

5. Complex sentences

The division of complex sentences into slash-units is a little more complicated.

(a) Divide a sentence at any coordinating conjunction (*and, but,* etc.; see Section 3.1.4 for more discussion.) as long as the following conjunct has a subject. If there is no subject, the second conjunct remains part of the first slash-unit.

```
A:  {D So, }  I'm a licensed teacher / {C but } jobs being what they are, I
substitute  teach -- /

A: Yeah, / {C and } we got him when he was about eight weeks old / {C and } {F
uh, } he's pretty okay, / we like him. /

A: -- {C and } I went / {C and } I signed up for class because then again,
you [ get, + {D you know, } get ] someone to guide you / {C and } you get the
tools you don't --
B:  #Yeah.# /
A:  -- feel like buying. /

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

B: {C But } they have to come in at [ a, + ] just the right angle. / If you
come in at the wrong angle, what happens is they bounce off and go back into
space -- /

A: -- that, {F uh, } it becomes more evident.  / Perhaps things that we didn't
think of before and just concentrated on the lawmaking or the results that
would be seen in public works or bills that are passed or, {D you know, } et
cetera like that -- /

A: -- {F uh, } [[ d-, + do, ] + do ] not worm their way around and influence
other people or perhaps lead other people astray. /
```

(b) *Because* introducing a causal clause and all other subordinate clauses are contained in the same slash-unit as the matrix clause.

```
B: -- he's indoor and all declawed because I know the, {F uh, } average
life span of an outdoor cat is eighteen months. /

A:  {D Well, } yeah, / [ I, + I'm ] signed up for a class in pottery
making <laughter> because I've never tried that yet -- /

A: [ I'm not, + I'm not ] in there, / I'm not one of those, {F uh, } {D you
know, } volunteers that's down at headquarters and whatever, {F uh, }
although I admire the people that do that / {C and } they certainly do get
inside information -- /
```

```
A: {C And } he usually is good about staying within them, although our next
door neighbors have a dog, too / {C and, } {F uh, } [ she, + she ] is good
friends with my dog. /

A:  He'd probably look pretty funny if we did that.  /

B:  I'd be afraid to walk around if I was that dog <laughter>. /

A:  I went there, /  we have Home Depot out here.  / I don't know if
you've heard of it. /

A:  Yeah, / {D well, } a friend of mine at work here said that he tried it
with his dog /

A: ... {C And } I've been fortunate that I can afford to work part-time. /

A: {C And } I saw a silver necklace that had these little teeny weeny
silver spatulas ... /
```

There are some cases in which what appears to be a formally subordinate clause, normally a
because clause, stands alone as a slash-unit. Often (but not always), *because* in these cases begins
a new turn and is not particularly associated with any main clause. See Section 3.1.4 for more
detail. In general in these cases, *because* is not causal, but rather means something like *The reason
I'm saying this is.*

```
A: That's probably a good idea.  / They're good though.  / {F Uh, } have you
ever had a Retriever before? /
B:  No,  / {C but } my husband had one when he was growing up.  /
A: {C Because } [ I, + I ] find that they're good around other animals like
cats and other dogs.  / They don't tend to get real jealous or territorial.  /
They like to play. /

A: That's right <laughter>. /  {D Well, } do you think we should continue to
support it at the level, / {D well, } I thought probably one of the things they
were doing that they didn't tell us was spying.  / {C Because } --
B:  Uh-huh.   /
A: -- I think those satellites, {D you know, } or not satellites, but the
spaceflights could really spy /
```

3.2.1   Turn taking

When a slash-unit is left unfinished at the end of one turn, it can only be completed by the SAME SPEAKER
(apparent completions by the other speaker are ignored). Any utterance by the other speaker in between
counts as a slash-unit on its own.

```
B:  <Laughter> I thought it was kind of a strange topic about corruption
in the government and --
A:  Yeah.  /
```

```
    B:  -- {F uh, } how many people are self serving.  /

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

If the speaker doesn't complete the slash-unit in their next turn, it counts as an incomplete slash-unit.

```
    A: {F um }.  What kind of -- -/
    B: [ [ It's, + {F uh, }
    A: -- pine? /
    B: It's, ] + {F uh, } it's, ] {F uh, } plywood, {F uh, } face, I guess.  /
```

This is the case even if the other speaker appears to finish the turn.

```
    A:  we did get the wedge cut out by building some kind of -/
    B:  A cradle for it. /
    A:  Yeah. / (MADE UP EXAMPLE)
```

3.2.2 Incomplete slash-units

When a turn (and its possible continuations, see Section 3.2.1) does not constitute a complete constituent, it is marked as incomplete with the symbol -/ . Since it is possible for a speaker to continue over more than one turn, it is necessary to check if the utterance is continued in the next turn before annotating a slash-unit as incomplete. (See also Section 3.3).

```
    A:  {F um }.  What kind of -- -/
    B:  [ [ It's, + {F uh, }
    A:  -- pine? /
    B: It's, ] + {F uh, } it's, ] {F uh, } plywood, {F uh, } face, I guess.  /

    A: ... I don't think they're always necessary. / If you put enough patience
    into, -/
    B:  Yeah, / just be consistent and diligent --
    A:  Uh-huh.  /
    B:  -- with it  / {C and, } {F um, } -/

    A:  -- unless things are so absolutely out of whack that, {F uh, } -/
    <end of file>

    A: We've had him for, -/ {D let's see, } he just had his fourth birthday. /

    B:  I happen to live not too far away from, -/ {F uh, } {D Well, } I've
    actually worked for the company that, {F uh, } has been blamed for the
    Challenger disaster. {F uh } --
    A:  {F Oh, } is that right?  /
```

```
B:  -- Morton Thiakol.  /

A:  -- {C but } you could see where they would come off in the, -/ how hot
it got because it was still the scorch on them which was quite
interesting. /
```

A dangling filler taking up a turn on its own (without *and* or anything) counts as a slash-unit, but dangling coordinating conjunctions and discourse markers (including *well, so, I mean*) for which there is no continuation in the next turn are counted as incomplete slash-units, except for *you know* which is a special case. See Section 3.1.3 for how to deal with *you know*. Non-speech sounds (<laughter>, <breathing>, etc.) at the end of an utterance should be included inside the slash-boundary. When they stand alone as a turn, they are just ignored and not counted as slash-units.

```
A:  he's pretty good. / He stays out of the street / {C and, } {F uh, } if I
catch him I call him / {C and }he comes back. / {D So }  [ he, + he's  ]  pretty
good about taking to commands [ and + --
B:  {F um. } /
A:  -- and  ] things.  /

B:  -- with it /  {C and, } {F um, } -/
A:  [ It's, + {F uh, } is  ] your cat an indoor cat or an --
B:  Yeah -- /
A:  -- outdoor cat?  /

B:  Right. / I've seen them / {C and then } the watermelon, the red [ with
the, + with the ] seeds painted in  / {C and, }  -/
A:  {F Oh, } yeah. / Uh-huh. /
B:  That's neat.  /

B:  I have a sister that lives in Dallas,  / that's how we got hooked up to the
program was [ her + --
A:  I see.   /
B:  -- her ] husband works for Penney's there.  /
A:  I see.   /
B:  {D So. }  -/
A:  Carolyn Rash? /
B:  Pardon me?   /

B:  -- {C and } if they come into it at a different angle there is so much
friction that it burns them up. /
A:  Right.   /
B:  {D So, } -/
A:  {D Well, } I saw [ one of, + some of ] those early space capsules at
Smithsonian.  / I'm sure you've --
B:  Oh.  /
A:  -- seen more of them than I have,  /

A:  Is her name Rash, Carolyn Rash?  /
B:  No, [ her name is, + {F uh, } my last name is ] Skidmore.
A:  I see.   /
```

```
B:  {C And } [ [ her, + her ] husband's name is, + his last name is ] Weight.
A:  I see.  /
B:  {C And, } {F uh, } -/
A: {D Well, } you probably know a lot about it. / I wasn't sure that they were
blaming that [ sp-, + space ] disaster on one company. /
B:  <Breathing>.
A:  I hadn't heard that.   /

A: #{D Well, } # yeah, / [ I, + I'm ] signed up for a class in pottery making
<laughter> because I've never tried that yet -- /
B:  {F Huh. } # {D So } you# --
A:  -- # {C but, }# -/
B:  -- have a M F A [ in, + in ] what #division of art?# /
A: # {D Well, } [ I, + I ] # actually have a Masters of art / {C and } it's in
art education. /

A: -- {C and } that way I have time for my kids too.  / {C And } I've been
fortunate that I can afford to work part-time.  / I don't know how long that's
going to last <laughter>. /
```

Note also that what might count as a slash-unit if it stands alone as a turn might count as incomplete if the speaker continues speaking, but has obviously begun a new slash-unit.

```
A: {C And } when it comes to being alone, {F uh, } -/ now if you give him the
freedom to walk around and go wherever he wants, he likes that, /
```

Subject pronouns alone always count as incomplete slash-units (despite arguably consisting of complete constituents).

```
B:  -- it was very unfortunate thing that
occurred there /  it's, -/
A:  Where do you live?  /
B:  we live in Utah. /

B:  {E I mean, } you -- -/
A:  Uh-huh <<Faint>>. /
B: -- [ I, + I ] think people tend to think that the government is not
corrupt <laughter>. /
```

3.3 Restarts

This section is based directly on Liz Shriberg's dissertation (Berkeley, 1994).

Restarts have the following form:

```
Show me flights from Boston on       uh      from Denver    on   Monday
                    |-----‘‘RM’’-------|--‘‘IM’’-|-----‘‘RR’’-----|
                                    ‘‘IP’’


‘‘RM’’ = reparandum
‘‘IP’’ = interruption point
‘‘IM’’ = interregnum (editing phase:  filled pause or editing terms)
‘‘RR’’ = repair
```

The entire restart with its repair is contained in square brackets. The IP is marked by a "+", and any IM material (filled pauses, etc.) follows the "+". A restart will "repaired" in the cleaned-up version by deleting the material between the open bracket and the IP (+).

```
Show me flights [ from Boston on  +  {F uh }  from Denver on ] Monday

                |-----‘‘RM’’------|--‘‘IM’’-|-----‘‘RR’’-----|
                                ‘‘IP’’
```

3.3.1 Simple restarts

- Repetition

    - simple repetition
        ```
        {C but } [ I, + I ] do know some of these places were doing that

        {F Oh, } [ I think, + I think ] two or three minutes is
        ```

    - repetition and insertion: at least one item is repeated on the right side, plus something new is inserted or added
        ```
        {E I mean } [ her, + all her ] kids were sick

        saying now [ he, + he's ] sick

        people would [ dr-, + have a drink ] of soda,

        [ Do you, + by mistakes, do you ] mean just
        like honest

        {C but } [ we're + at the same time we're ] real scare
        d about
        ```

```
          So [ it work-, + it did end up working ] out pretty well?
```

  – repetition and deletion: at least one item is repeated on the right side, plus something is deleted
```
          [ I just, + I ] noticed it

          [ I really like + I like ] pizza
```

  – repetition and substitution: at least one item is repeated on the right side, plus there is a substitution of another element on the left, by a different one on the right
```
          [ why didn't he, + why didn't she ] keep him home with her?

          {F Uh, } [ I know, + {F uh, } I believe ] it was last year

          [ you would know, + the whole world would know ] what you we
          re facing. /

          the judge gave the woman who [ shot her, + shot the girl ]

          it was difficult to [ find someone, + {F uh, } find a panel ] that would
```

- Substitution

  Substitution plus repetition was covered in the previous section. If the left side has more than one word, at least part of it must be repeated on the right side for the whole thing to count as a restart (as in the cases under repetition plus substitution).

  Substitution without repetition is only possible if the left side contains a single word. Something, however, must carry over from the left to the right. In the case of single words this will most often be the part-of-speech.

```
          ARTICLE: I gave her [ a + the ] dog

          PRONOUN: [ he + she ] reported it

          PREPOSITION: I put it [ in + on ] the table

          ADJECTIVE: a [ blue + green ] block

          NOUN: a blue [ triangle + square ]

          VERB: I [ liked + loved ] the opera

          ETC...
```

  This means that you cannot make things like *I gave her the, uh, what she wanted* restarts because *the* and *what* are not the same part of speech. Instead this type of example will fall under **Deletion**; see below.

It is possible for a single word on the left to be replace by a phrase on the right, as long as the phrase on the right fulfills the same function as the word on the left.

```
a [ blue + kind of green ] block

[ she + my mother ] bought me a present

they gave us [ breakfast + a snack ]
```

- Deletion

  Deletion is used when the section that needs to be removed is not replaced by anything, according to the rules given above.

  There are two ways to delete material without replacing it:

  – a restart without repair

  ```
  B: [ [ Did they, + did they, ] + like on bottles, did they ] give you [ a
  + ] so many cents back

  A.21: Yeah, / I do <laughter>. / Yes, / {F uh, } I don't work, though, / {C but
  } I used to work [ and, + ] when I had two children. /

  A.61: -- {C but, } {F uh, } seems like, {D you know, } -/ goll-, -/ what if he
  crie d at the wrong time / <laughter> {C or } the phone rang [ or, + ] {D you
  know, } at the wrong time <laughter>. /

  A.53: <Breathing> {D Well, } their knowledge <breathing>, with, {F uh, } better
  communication processes in this world now that's reached them is more educated
  [ than + ] to the, {F uh, } ways of the western world, /

  B: {D You know, } all kinds of bottles [ [ they, + they, ] + they ] really
  charge people [ to, + ] I guess when you purchase them [ {C and, } + {C and
  then } ] when you turn them back in. /
  ```

  – an incomplete slash-unit

  ```
  A: The rec-, -/ we live in Plano /

  B:  -- {C but } ] [ it's, + {F uh, } it's ] quite, -/ they say that the green
  glass now, that there's a big glut --

  B: [ It, + it's, ] -/ people just, -/ [ they, + they ] really don't, /

  B.22: {F Oh, } yeah, / yeah, / it's a-, -/ the benefit of the doubt to the last
  iota is, {F uh, } based on the, {F uh, } person who is accused. /

  B.36: -- it's also, -/ I used to live in Georgia /
  ```

  At the beginning of a unit, ALWAYS use an incomplete slash-unit, in the middle ALWAYS use a restart without repair

3.3.2 Complex restarts

Complex restarts are those with more than one interruption point (+). There are two types of complex restarts: (1) chaining and (2) nesting.

- Chaining

  In chaining, something from each part carries over into the next one.

  - simple repetition

    ```
    [ [ I + I ] + I ] don't want it
    ```

  - simple repetition with insertion

    ```
    [ [ the + the boy ] + the big boy ]
    ```

  - simple repetition with deletion

    ```
    [ [ I really + I ] + I ] like it
    ```

  - Single words can be repeated or substituted (following the rules under **Simple restarts**)

    ```
    [ [ he + she ] + she ]
    ```

    ```
    [ [ he + she ] + they ]
    ```

  - Multi-word complex restarts are more complicated. The rule to keep in mind is that if there is any repetition, it MUST BE BETWEEN CONTIGUOUS PARTS OF THE RESTART. So in the following example, *flights* repeats between part 1 and part 2, and *Delta* repeats between part 2 and part 3.

    ```
    [ [ the flights + Delta flights ] + Delta fares ]
          part1             part2               part3
    ```

    An example like the following however, is not a chain.

    ```
    I liked, uh, I, I liked it a lot
    ```

    Note that a chain would look as follows:

    ```
    [ I liked, + {F uh,}  I, ] + I liked ] it a lot (THIS IS BRACKETED WRONG)
       part1               part2       part3
    ```

    Here, although *I* repeats in all three parts, *liked* repeats from part 1 to part 3; that is across NON-CONTIGUOUS PARTS. This is what is called **nesting** and how to bracket it is discussed in the next section.

- Nesting

  Nesting is the case when you have a restart within a restart.

  ```
  [ I liked, + {F uh, } [ I, + I ] liked ] it a lot
                        part1 part2              INNER RESTART
     part1        -------------part2------          OUTER RESTART
  ```

Here, in order to get *liked* into contiguous parts of the restart, *I, I* within the second half of the first restart is made a restart of it's own. Imagine that restarts are resolved from the inside out and so the internal restart *[I + I]* is resolved first leaving only *I* . The outer restart now has the form *[I liked, + F uh, I liked]* which is a perfectly well-formed repetition.

Here are some examples:

```
[ I, + I ] think the line item veto [ is [ not, + not, ] + is not ]
necessarily a bad thing,

It's, must be depressing [ to, + to ] [ walk, + walk ] the halls [ and see, + [
an-, + and ] see ] all these other people who really don't know where they are.
/

{F Uh, } [ I, + before [ I, + I, ] ] {F uh, } <lipsmack> {F uh, } got my degree
-- -/

I'm sure [ that the, + that [ the, + {F uh, } the ] ] staff [ l-, + learn ]
what's normal for this person

[ when I, + when I ] see that money taken out of my paycheck [ [ each, + each ]
week, I, + {E or } [ each, + every ] other week, [ I, + I ] ] really think that
money's history

B.62: Yeah. / [ [ I've, + I've heard some, ] + {E I mean } I've heard [ [ sta-,
+ sta-, ] + statistics ] ] and things

{C and } [ I'm actually making just about as much as, + {D you know, } I'm [
m-, + making ] actually [ over average, + over the average ] ] --
```