

# **Quick Rich Transcription (QRTR) Specification for English Broadcast Data**

(XTrans-Format Version)

Version 2.0 – December 4, 2008

Linguistic Data Consortium

<http://projects ldc.upenn.edu/gale/Transcription>

1	Introduction.....	3
2	Data.....	3
3	Segmentation Task.....	4
3.1	Introduction.....	4
3.2	Timestamping the Audio.....	4
3.3	What to Segment.....	5
3.4	Overlapping and Simultaneous Speech.....	5
4	Sentence Units (SU).....	6
4.1.1	Statement SUs.....	6
4.1.2	Question SUs.....	7
4.1.3	Incomplete SUs.....	7
4.1.4	Recognizing SU Boundaries.....	8
5	Identifying Section Boundaries.....	8
6	Speaker Identification.....	9
6.1	Speaker Type.....	9
6.2	Names and Identifiers.....	10
6.3	Native and Non-native Speakers.....	10
7	Transcription.....	11
7.1	Orthography and Spelling.....	11
7.1.1	Spelling.....	11
7.1.2	Punctuation.....	11
7.1.3	Numbers.....	11
7.1.4	Proper nouns.....	11
7.1.5	Contractions.....	12
7.1.6	Acronyms.....	12
7.1.7	Spoken letters.....	12
7.2	Disfluent Speech.....	13
7.2.1	Filled Pauses and Hesitation Sounds.....	13
7.2.2	Partial Words.....	13
7.2.3	Mispronounced words.....	13
7.2.4	Idiosyncratic words.....	14
7.3	Speaker Errors and Non-standard Usage.....	14
7.4	Foreign Languages and Dialects.....	14
7.4.1	Foreign Languages.....	14
7.5	Noise.....	14
7.5.1	Background Noise.....	14
7.5.2	Speaker Noise.....	14
7.6	Hard-to-understand Regions.....	15
7.7	Summary of rules.....	15
8	English Interjections.....	15
9	Summary of Conventions.....	2

# 1 Introduction

The goal of quick rich transcription (QRTR) for broadcast news and broadcast conversation is to produce a verbatim, time-aligned transcript with minimal but useful markup. QRTR also identifies some salient structural features of the broadcast and provides speaker identification.

The elements of a quick rich transcript include:

- verbatim transcription
- time-aligned section boundaries, speaker turns and sentences (segmentation)
- section and sentence type identification
- speaker identification
- standard treatment of common spoken phenomena

Transcription begins with audio segmentation. This involves "timestamping" structural boundaries including sections (story transitions and the like), speaker turns and sentences. Speakers are identified by name where possible, or by a unique identifier, such as "*speaker1*". Once audio has been virtually segmented into smaller units, annotators transcribe the content of each segment. Special conventions are used to flag certain speech phenomena like disfluencies and mispronounced words. Quality control checks verify the format of the resulting file as well as overall transcript quality.

On the scale of less-to-more careful transcription, QRTR falls somewhere in the middle. It differs from Quick Transcription (QTR) in that each sentence unit is manually timestamped and labeled for its type, such as *statement*, *question*, or *incomplete statement*. QRTR differs from careful transcription (CTR) in the amount of detail in the transcript markup, the number of features identified, the degree of accuracy and completeness of the transcript, the amount of time taken to complete the file, and the number of quality checks that are performed on the finished product.

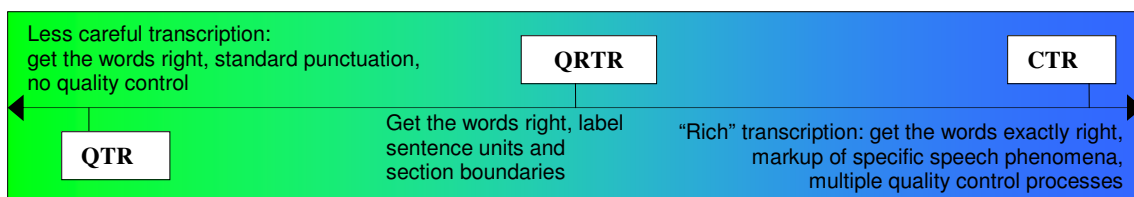


Figure 1: Range of detail in transcription tasks at LDC

## 2 Data

These guidelines pertain to data in the following genres:

- *Broadcast News (BN)* consisting of "talking head"-style news broadcasts from radio and/or television networks.

- *Broadcast Conversation (BC)* consisting of talk shows, interviews, roundtable discussions and other interactive-style broadcasts from radio and/or television networks.

Data is divided into files, which typically correspond to a recording of one broadcast from a single program. Files are typically 30 to 60 minutes in duration, though they may be of any length. Files come from a range of radio, television, satellite and web broadcast sources from around the world. Each show is pre-designated as BN or BC based on its characteristic content. Note however that BN shows can sometimes contain stories that are conversational, while BC shows can include hard news reports.

### **3 Segmentation Task**

#### **3.1 Introduction**

Transcription begins with segmentation. During the segmentation task, annotators virtually chop an audio recording into smaller units that correspond to certain features of the broadcast, for instance sentences or speaker turns. Each segment must be timestamped, that is, time-aligned with the audio to identify where the segment starts. Segments are also classified by type and subtype. We identify three kinds of segments in the QRTR task: Sections, Turns, and Sentence Units. These are arranged hierarchically (sections contain turns, turns contain sentences).

Annotators begin segmentation by identifying sentence units, labeling speaker turns as they go along and noting new section boundaries when they occur.

Segmentation can be done with the keyboard only, with the mouse only, or with a combination of both. After you've become familiar with basic XTrans functionality, you will find that using only the keyboard is both faster and more intuitive than using the mouse.

#### **3.2 Timestamping the Audio**

Timestamps are required for each segment and mark the start and end times of each segment. Timestamps are designated in seconds, rounded to the nearest thousandth of a second. XTrans does not show timestamps in the transcript display, but they are properly encoded in the output file. A timestamp also may be called a segment boundary.

Because broadcast speech is recorded on a single audio channel, segment boundaries occur one right after the other, in direct succession and typically without intervening periods of unsegmented audio. Small gaps between segments should indicate a region of non-speech, like music, sound effects, background noise, or some other untranscribed event. Continuous speech from a single speaker that has no pauses or other non-speech events should not contain gaps between segments. (Using the mouse sometimes is likely to leave

small gaps within continuous speech, while using the keyboard may avoid the occurrence of small gaps.)

Segment boundaries must be placed in between words, not inside of them or at the very edges of words where speech sounds could be truncated. It is critical that the time and the audio are perfectly aligned, so that the words transcribed within each timestamp match the speech associated with that timestamp.

Annotators should take care not to cut off the beginnings or ends of sounds or words when inserting segment boundaries. Good places to insert timestamps are during pauses, breaths or other non-speech events, which typically occur at sentence unit (SU) boundaries.

### 3.3 What to Segment

All speech must be segmented and classified into sections (news reports, conversational segments or non-news). News reports and conversational segments must also be segmented into SUs, with speakerIDs added. Non-news sections should **not** be segmented into smaller units or labeled for speakerID, and they should not be transcribed.

Very brief (under 0.5 seconds) periods of silence, music, background noise or other types of non-speech that occur while someone is speaking should simply be included within that segment, or split between two adjoining segments. No other treatment is necessary. Lengthy segments of non-speech that interrupt a speaker's turn or come between speaker turns should be excluded from segmentation entirely. For long pauses within a SU segment, do not split the SU into two units, but rather maintain the intactness of the SU segment.

### 3.4 Overlapping and Simultaneous Speech

In broadcast data, overlapping speech from two or more speakers is a relatively frequent occurrence. XTrans prevents users from creating overlapping segments for a single speaker. Although broadcast audio is single channel, XTrans creates a separate *virtual* channel for each unique speaker. There is then no prohibition on overlapping segments from two or more *different* speakers. Hence, transcription of overlapping or simultaneous speech is very straightforward. To create an overlapping speech segment, transcribers simply perform segmentation and speakerID normally. The image below shows an overlapping region as it appears in XTrans:

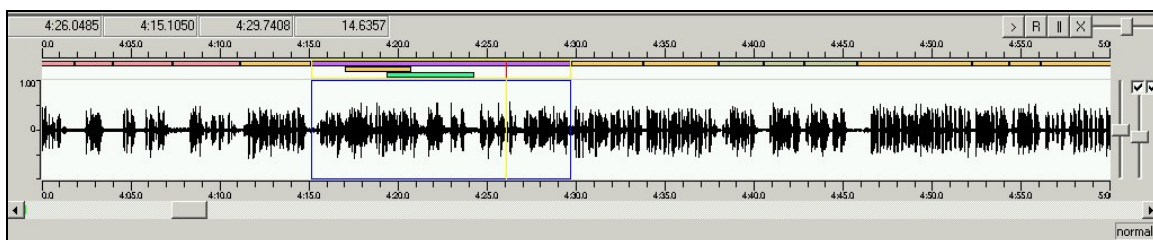


Figure 2: Overlapping speech region in XTrans

## 4 Sentence Units (SU)

Segmentation begins with identification of sentence unit boundaries in news reports and conversational segments. A sentence unit (SU) is a natural grouping of words produced by a single speaker. SUs have semantic cohesion – that is, they can have some inherent meaning when taken in isolation; and they have syntactic cohesion – that is, they have some grammatical structure<sup>1</sup>. In written language, sentences are usually designated by punctuation like periods or question marks. When creating SU boundaries for spoken language, our goal is to identify a semantically and syntactically cohesive group of words that constitute a reasonable sentence-like unit. For each SU, annotators identify its type and insert a timestamp at the start of the SU.

We distinguish three types of SUs: statements, questions and incomplete sentences. After identifying the boundaries of a SU and inserting a timestamp, annotators can simply use the tool to indicate the SU type of the segment. Generally, the SU segments should be coherent with the punctuations annotators use in transcription, as follows:

period	end-of-sentence markup for Statement SUs	.
question mark	end-of-sentence markup for Question SUs	?
double dash	end-of-sentence markup for Incomplete SUs	--

Annotators will note that standard punctuation typically includes commas as well. For purposes of the QRTR task, we do not identify an SU (or sub-SU) unit that corresponds to a comma. Commas may be added into transcripts for human readability, but it should be understood that the existence of a comma does not imply the existence of a sentence boundary. See Section 7.1.2 for a full discussion of use of punctuation in QRTR transcripts.

The sections that follow provide language-specific rules for identifying SUs of each type.

### 4.1.1 Statement SUs

Statements are declarative sentences or fragments, and are usually punctuated by a period or exclamation point. For instance,

On the other hand we do see the positive aspects of Arab society.

Today's topic is divorce.

Maybe she wants to get a master's or complete a doctorate.

A step in the right direction.

An important step.

---

<sup>1</sup> Note however that incomplete SUs may contain incomplete semantic and/or syntactic content.

A backchannel is a word or phrase that provides feedback to the dominant speaker, indicating that the non-dominant speaker is still paying attention to the conversation. In QRTR, backchannels are treated as statement SUs. When a speaker chains together several backchannels in succession, annotators tag them as one single statement. For instance,

```
speaker1: I enjoyed the concert last night.  
speaker2: Uh-huh uh-huh yeah.
```

#### 4.1.2 Question SUs

The question label should be used for a complete sentence that functions as an interrogative. The expected end-of-sentence punctuation for a question is a question mark.

```
Dr. Amin, are children more susceptible to trauma after the  
tsunami than adults?
```

A tag question is a phrase added to the end of an utterance that invites the listener to give feedback. Tag questions usually do not stand alone as a question, but rather form a complete question with the previous utterance:

```
You've been working there for years, right?
```

```
It has a cure, doesn't it?
```

Rhetorical questions should also receive a Question SU label:

```
Isn't it said that peace is always acceptable and there is no  
such thing as a good war?
```

#### 4.1.3 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 and is marked with a double dash.

Incomplete SUs frequently occur in two situations. When a speaker interrupts him/herself and then restructures the utterance and continues speaking on the same topic, an incomplete sentence exists. For instance:

```
I said to --  
I am not in agreement with this subject at all.
```

```
The only way is ensuring Iraqi unity with all its sects and and -  
-
```

The other type of incomplete SU occurs when one speaker's turn is cut short by an interruption from the other speaker, or when the speaker trails off at the end of his/her turn and abandons it completely, without then restructuring it or continuing along the same lines:

speaker1: Their children, their grandchildren, their great great grand- --  
speaker2: Dr Amin, let me interrupt you to introduce our guest from Cairo.

#### 4.1.4 Recognizing SU Boundaries

It can sometimes be difficult to determine where a sentence boundary exists between two clauses, and where those clauses belong within the same SU. Annotators rely primarily on the meaning conveyed by the utterance and apply SU breaks in accordance with the (mainly syntactic) rules described in the guidelines. However, annotators may sometimes rely on prosodic features like sentence intonation or pauses to determine where to place an SU boundary. In practice, SU boundaries tend to occur at the ends of fragments, simple sentences and complex sentences.

Complex sentence are very common in spoken English and can be tricky to segment into SUs. In general, annotators should lean toward creating a single SU for complex, multi-part sentences. This is particularly true when two parts (clauses) of the sentence depend on one another for the completion of an idea, for instance:

Not only do we pollute the water but we also let the sewers empty straight into the sea.

As far as emotional reactions some people are unable to sleep without a sleeping pill or a sedative.

A bomb exploded in a tunnel near the hotel and many cars in the area were damaged.

I followed as did many others the appearance of a division between Saudi and Bahrain and both members of the Gulf Cooperation Council.

## 5 Identifying Section Boundaries

QRTR also calls for identification of section boundaries. A section is a topically contiguous segment of the broadcast. Sections begin at SU boundaries. At the beginning of each new section, annotators simply insert the appropriate section label. Consecutive sections of the same type should receive separate section boundary labels, except in the case of consecutive commercials and other untranscribed segments which should be grouped together as a single section – or simply not segmented at all (see Section 3.3). All audio in a speech file must be assigned to a section.

We recognize three section types:

- **Reports** include typical "talking head" news broadcast, with an anchor reading the news. This may also include broadcasts from reporters in the field. News reports may be of any length, as long as they constitute a complete,

cohesive news report on a particular topic. Note that single news stories may discuss more than one related topic. When reports of similar content are adjacent to one another in a broadcast, it is often difficult to tell where one story ends and the next begins. Annotators are instructed to rely on audio cues (speaker changes, music, and pauses) to inform their judgments. When in doubt, do not create a new section boundary.

- **Conversations** include highly interactive segments of a broadcast, including roundtable discussions, interviews, call-in segments, debates and the like. Some conversation sections are quite long and can contain multiple topics. Annotators should create a new section boundary only at natural breaks in the flow of conversation, for instance, when there is a major shift in topic, or when a new panelist joins a roundtable discussion. If in doubt, the annotator should avoid creating a new conversation boundary.

It may sometimes be difficult to tell the difference between a report and a conversational segment. When in doubt, annotators should use the label report.

- **Non-news** text includes segments like commercials, station identifications, public service announcements, promotions for upcoming shows and long musical interludes. **Note that non-news sections are not transcribed or further annotated in any way (including speaker ID or SU segmentation). In fact, it is not necessary to segment these regions at all.** (See Section 3.3)

However, if a non-news section ***is*** segmented, once it has been identified and labeled, it should be ignored for the rest of the transcription task. If multiple non-news sections follow one another within a transcript, they should be grouped together as a single section. This is different from multiple consecutive news or conversational reports, which should be separated into multiple sections.

## 6 Speaker Identification

In addition to identifying SUs and section boundaries, annotators also label the identity of speakers within a broadcast. Speaker IDs are required with each SU segment<sup>2</sup>. Each speaker label has three elements: speaker type (required), non-native status (optional) and speaker "name" if available.

### 6.1 Speaker Type

All speakers must be assigned a speaker type. There are four speaker types as follows:

- Female – used for adult females
- Male – used for adult males
- Child – used for children of either sex
- Other – used for speakers in unison, non-human (computer) voices, altered voices, unknown speaker sex, etc.

---

<sup>2</sup> The XTrans toolkit requires annotators to provide speaker ID for each SU annotation.

## 6.2 Names and Identifiers

All speakers must be identified by name. When the name is not known, annotators use a **unique** identifier for each speaker (speaker1, speaker2, etc).

When names are known, they should be written out in full. For names with multiple spellings or transliterations, the most common variant should be used. If in common practice the name contains a middle initial or an appositive like "Jr.", these should be included and spelled out in full. All names must be written using the most common transliteration. Capitalization should follow standard conventions.

The spelling of speaker IDs must be consistent within a broadcast file, and wherever feasible across different broadcast files as well. It is also important that the spelling of names within a transcript match the spelling of the name in within the speaker ID label. For instance, if the transcript uses the transliteration "Osama bin Laden", then the speaker ID should also use "Osama", not "Usama". When a speaker is not identified by name within a recording, the speaker should be labeled with a unique numerical identifier, e.g. *speaker14*. Each anonymous speaker is assigned a unique number that should be used for every instance of that speaker throughout the broadcast. Anonymous speaker IDs cannot be re-used for different speakers in the same file, regardless of gender or speaker type.

Following are some examples of standard speakerIDs:

Acceptable	Unacceptable
George_W_Bush	George W. Bush
Mary_O_Malley	Mary O'Malley
Abu_Musab_al_Zarqawi	al-Zarqawi
Condaleeza_Rice	Condi
Hank_Williams_Junior	Hank Williams, Jr.
speaker12	Unknown speaker # 12

## 6.3 Native and Non-native Speakers

In addition to labeling speaker type and name, annotators also indicate when a speaker is non-native; that is, when they use a language variety other than the target, or when they speak the target language with a discernable foreign accent. Targets for the current task are

- Arabic – Modern Standard Arabic (MSA)
- Chinese – Mainland Mandarin Chinese
- English – American English

Speakers using other varieties/dialects of these languages, or speaking these languages with a heavy non-standard accent (for instance, Cantonese-accented Mandarin, or British English) should be marked as non-native.

## 7 Transcription

Quick-rich transcription requires annotators to produce a verbatim transcript of all speech within a file and to add minimal markup to capture salient features of the speech. Standard writing conventions, including spelling and punctuation, are used for ease of comprehension and readability. Transcripts must be produced in UTF-8 (Unicode) encoding. Transcripts should be spell-checked for common misspellings or typographical errors before they are considered finished.

### 7.1 Orthography and Spelling

#### 7.1.1 Spelling

Transcribers should use standard orthography and word spelling. All files must be checked for typos and misspellings after transcription is complete. When in doubt about the spelling of a word or name, annotators consult a standard reference, like an online or paper dictionary, world atlas or news website.

#### 7.1.2 Punctuation

Annotators should use standard punctuation for ease of transcription and reading. Acceptable punctuation is limited to the following:

Type	Usage	Symbol
period	end-of-sentence markup for Statement SUs	.
question mark	end-of-sentence markup for Question SUs	?
double dash	end-of-sentence markup for Incomplete SUs	--
comma	sentence-internal, used to aid readability	,

Transcripts should not contain quotation marks, exclamation marks, colons, semicolons, single (stand-alone) dashes, or ellipses in transcribing. Punctuation should be written as it normally appears in standard writing, with no additional spaces around the punctuation marks.

#### 7.1.3 Numbers

All numerals should be written out as complete words instead of numbers. Hyphenation for numbers twenty-one through ninety-nine is optional. Here are some examples of properly formatted numbers.

My number is two one five five five five oh oh seventy-two

That'll be nine ninety-five, plus tax is ten seventy-one.

Listen to ~CNN two at eleven thirty.

#### 7.1.4 Proper nouns

No special markup is provided for proper nouns. Note however that spelling of names within a transcript should match the spelling of the name in within the assigned speaker ID. For instance, if the speaker ID uses the transliteration

"Osama bin Laden" the transcript should also use "Osama" when that name is spoken, not "Usama" or some other form.

### 7.1.5 Contractions

Contractions are transcribed exactly as the speaker says them. The table below, while not comprehensive, shows some examples of how to transcribe common contractions.

Complete Form	Spoken As	Transcribed As	Incorrect
I have	I've	I've	
Cannot	can't	can't	
Will not	won't	won't	
you have	you've	you've	
Could not	couldn't	couldn't	
Should have	should've	should've	<b>should of, shoulda</b>
would have	would've	would've	<b>would of, woulda</b>
It is	it's	it's	<b>its</b>
its (possessive)	Its	its	<b>it's</b>
Marvin (possessive)	Marvin's	Marvin's	
Marvin is	Marvin's	Marvin's	
Marvin has	Marvin's	Marvin's	
<b>Going to</b>	<b>gonna</b>	<b>going to</b>	<b>gonna</b>
<b>Want to</b>	<b>wanna</b>	<b>want to</b>	<b>wanna</b>
<b>Got to</b>	<b>gotta</b>	<b>got to</b>	<b>gotta</b>
<b>Because</b>	<b>cuz</b>	<b>because</b>	<b>cuz, cos, 'cause</b>

### 7.1.6 Acronyms

Acronyms that are pronounced as a single word should be written in all capital letters, and preceded by the @ symbol:

@NASA  
@UNICEF

### 7.1.7 Spoken letters

Spoken letters and abbreviations that are pronounced as a sequence of individual letters should be written as a string of letters with no spaces between them, preceded by a tilde (~). For instance:

~RESPECT, find out what it means to me.

I would have gotten a ~B if I hadn't fallen asleep.

You're watching ~NBC news.

## 7.2 Disfluent Speech

Regions of disfluent speech are particularly difficult to transcribe. Speakers may stumble over their words, repeat themselves, utter partial words, restart phrases or sentences, and use hesitation sounds. Annotators should attempt to accurately transcribe as many of these utterances as possible. For purposes of QRTR, annotators should not spend too much time trying to precisely capture difficult sections of disfluent speech, but should make their best effort to transcribe what they hear after listening to the segment once or twice, and then move on.

### 7.2.1 Filled Pauses and Hesitation Sounds

Filled pauses are non-lexemes (non-words) that speakers employ to indicate hesitation or to maintain control of a conversation while thinking of what to say next. The spelling of filled pauses is not altered to reflect how the speaker pronounces the word (e.g., typing AH for a loud "ah" or ummmm for a long "um".) There is a restricted set of filled pauses for each language, with established spelling conventions. For English, filled pauses are limited to the following:

ah eh er uh um

These words are NOT capitalized. If a filled pause occurs at the beginning of a sentence, write it with lower-case letters and capitalize the first *content word* of the sentence, for example:

um **What** do you think of today's topic?

### 7.2.2 Partial Words

When a speaker breaks off in the middle of the word, annotators transcribe as much of the word as can be made out. A single dash - is used to indicate point at which word was broken off.

It is continu- continuing.

Wh- what do you mean?

### 7.2.3 Mispronounced words

A plus symbol + is used for obviously mispronounced words (not regional or non-standard dialect pronunciation). Annotators should transcribe using the standard spelling and should not try to represent the pronunciation. Annotators should not miss this mark up whenever there are any mispronounced words and should not transcribe the mispronunciation.

He'll +probably -- I mean probably go with me tomorrow.

Keep in mind that this symbol should only be used for obviously mispronounced words – words that the speaker himself or herself would consider to be incorrect.

Dialect pronunciations or other common variants of words should not be marked as mispronunciations.

### 7.2.4 Idiosyncratic words

Occasionally a speaker will make up a new word on the spot. These are not the same as slang words, but rather are words that are unique to the speaker in that conversation. If annotators encounter an idiosyncratic word, they should transcribe it to the best of their ability and mark it with an asterisk \*. For instance,

Do you dress like a \*schlump yet?

Why she said \*drr I don't know

### 7.3 Speaker Errors and Non-standard Usage

Annotators should not correct grammatical errors, e.g. "I seen him" for "I saw him" should be transcribed as spoken. The same goes for misused words and non-standard usage, e.g.

Annotators should transcribe *what is spoken*, not what they expect to hear.

### 7.4 Foreign Languages and Dialects

#### 7.4.1 Foreign Languages

Portions of speech in any language other than the target language are annotated using the `<foreign lang="LANGUAGE"> text </foreign>` convention to indicate the language and to transcribe the words that are spoken in that language if annotators know the language, for instance:

After four days, Jean-Luc declared, `<foreign lang="French"> je suis prêt à faire la paix </foreign>`. I'm ready to make peace.

If the annotator does not know the name of the language or what is being said, they should mark the language as "unknown".

`<foreign lang="unknown"> </foreign>`

Note that borrowings from other languages, such as *rendezvous* or *siesta* are not marked as foreign words.

### 7.5 Noise

#### 7.5.1 Background Noise

Transcribers are not required to specially label background noise or sound effects. Note however the convention for indicating long periods of non-speech within or outside a SU segment (Section 3.3).

#### 7.5.2 Speaker Noise

Speaker-produced noise is identified with one of the following four tags:

```
{laugh}  
{cough}  
{sneeze}  
{lipsmack}
```

However, speaker noise annotation is not required for QRTR.

## 7.6 Hard-to-understand Regions

Sometimes an audio file will contain a section of speech that is difficult or impossible to understand. In these cases, annotators should use double parentheses (( )) to mark the region of difficulty. It may be possible to take a guess about the speaker's words. In these cases, annotators transcribe what they think they hear and surround the area of uncertain transcription with double parentheses:

```
The new policy will allow us to ((move forward)).
```

If an annotator is truly mystified and can't at all make out what the speaker is saying, s/he uses empty double parentheses to surround the untranscribed region. Do NOT skip the region.

## 7.7 Summary of rules

1. **Transcribe what you hear, not what you think is correct.**
2. Do not add in your own words if they are not in the audio or drop out words even it is not grammatical.
3. No normalization of dialectal words.
4. Do not transcribe accent features. Use standard orthography.
5. Do not skip words that are hard to understand. Use (()).

## 8 English Interjections

This list, while not complete, shows the standardized spelling of some English interjections:

ach	huh	okay	whoops
duh	huh-uh	oof	woo-hoo
eee	jeepers	ooh	wow
ew	jeez	uh-huh	yay
ha	mhm	uh-oh	yeah
hee	mm	uh-uh	yep
hm	mm-mm	whew	yup
hm-mm	nah	whoa	

## 9 Summary of Conventions

Category	Condition	Markup	Example	Explanation
Orthography and spelling	Numbers	Spelled out	twenty-five, one oh nine, one hundred thirty-seven	Write out in full; hyphens optional for twenty-one through ninety-nine
	Standard contractions	Transcribe as spoken.	can't, I'm	If you hear a contraction used, write it as a contracted form.
	Non-standard contractions	Not used	going to, want to	Do not use non-standard contractions. Write the words out in full.
	Punctuation	Comma, question mark, period, double dash	, ? . --	Limited to these four symbols.
	Pronounced acronyms	@	@NAFTA	Write letters with all caps, no space between letters.
	Individual letters	~	~I before ~E ~YMCA	Individual letters spelled out, capitalized, each with ~ Strings of letters in all caps, no space between letters, string marked with ~
Disfluent speech	Filled pauses	No markup	ah, eh, er, uh, um	Limited to small list for each language; use standardized spellings
	Partial words	-	absolu-	Speaker-produced partial words are indicated with a dash. Transcribe as much of the word as you hear.
	Speaker restart	--	I think -- I thought he was there.	Used when the speaker stops short and then repeats themselves or abandons the utterance completely, restarting with a new sentence.
	Mispronounced words	+	+probably	Mispronounced word (a speech error). <b>NOTE:</b> Do not use this symbol to indicate non-standard but common regional/social dialect pronunciations. Transcribe non-standard pronunciation variants or mispronounced words using standard orthography.
Noise conditions	Speaker noise	{ }	{cough} {laugh} {sneeze} {lipsmack}	Sounds made by the talker. Limited to these four. <b><u>NOT required markup for QRTR</u></b>
	Non-speaker noise	Not used		<b><u>NOT required markup for QRTR</u></b>
Other markup	Semi-intelligible speech	((text))	They lived ((next door to us))	This is the transcriber's best attempt at transcribing a difficult passage.
	Unintelligible speech	(( ))	(( ))	This indicates an entirely unintelligible passage.
	Idiosyncratic words	*	*poodleish	Speaker uses a "made-up" word. <b>NOTE:</b> Do not use for non-standard dialect terms or misused words.
	Foreign language	<language text> </language text>	<French> merci </French> <foreign>	This is used to indicate foreign speech. If the word is unknown, leave it out. If the language is unknown, merely write <foreign>. <b>NOTE:</b> Do not use this convention for foreign borrowings that are common in the target language, e.g. <i>apropos</i> .
	Proper names	no special markup	Secretary of State Condoleezza Rice	Check proper name spelling on news website or dictionary; names should be consistent within and across broadcast transcripts
	Interjections	no special markup	uh-huh, yeah, mhm	Use standardized spellings