TRANSCRIPTION GUIDELINES

Linguistic Data Consortium

http://www.ldc.upenn.edu/Projects/DASL/SLX

Version 2.0 – September 30, 2003

1 Introduction

This document specifies the guidelines followed in creating time-aligned transcripts of the SLX interview data. The overall goal of the transcription process is to provide an accurate, verbatim (word-for-word) representation of each recording in its entirety. The resulting transcript is further time-aligned with the audio file, and additional features of the audio signal and speech are identified using special markup.

The sections that follow contain the detailed instructions given to the annotators who were responsible for creating the time-aligned transcripts.

2 Segmentation

2.1 Overview

Segmentation involves "timestamping" the audio file for a given speaker, indicating structural boundaries like turns, utterances and phrases within the interview and allowing the finished transcript to be time-aligned with the corresponding audio file. Segment boundaries also make transcription of the audio easier, by allowing the annotator to listen to small chunks of segmented speech at a time.

The segmentation process begins with creation of initial timestamps that roughly correspond to speaker turns. This is followed by a second pass that refines the initial timestamps and creates additional breakpoints at turn- and utterance-internal boundaries like pause and breath groups.

2.2 Process and Guidelines

Using the Transcriber tool, annotators first create one set of timestamps, or segments, for each speaker in the corpus. Although most corpus users will focus their attention on the main subject(s) within each interview, our goal is to create a complete transcript of all speakers in order to provide context and to allow for careful examination of discourse structure and interview technique.

Annotators create a separate segmentation file for each speaker in an interview, rather than a single segmentation file combining all speakers. This approach allows segmentation to proceed more quickly and enhances the precision of the task by allowing the annotator to focus on the speech of a single speaker at a time. The Transcriber tool allows initial segmentation to proceed at close to real time. Within a segmentation file for a given speaker, annotators begin playing the audio then hit <enter> each time the target speaker begins speaking, and again when the speaker finishes his/her turn. This produces two kinds of segments: those that contain utterances for the target speaker and those that do not. Segments that do not contain the utterances of the target speaker will contain either silence, non-speaker noise or utterances by another speaker in the session. This second type of segment exists purely to delineate the segments of the target speaker. This approach is necessary because of the design of the Transcriber tool, which does not allow overlapping segments, insists that an audio file always be fully segmented, and assumes that the end of any one segment is the beginning of the next.

The initial segmentation pass creates a set of rough boundaries that map more or less to speaker turns. A second segmentation pass refines the initial timestamps to provide more fine-grained alignment. These finer timestamps, or breakpoints, are inserted to break up long speaker turns for ease of transcription. Annotators insert breakpoints around breath groups, at ends of sentences or phrases and at noticeable pauses. Breakpoints are also inserted around lengthy

(greater than 0.5 seconds) non-speech events within a speaker's turn. These include things like background noise and silence. During the second pass annotators also correct any errors in the original segmentation.

Because breakpoints are inserted primarily for ease of transcription, their exact implementation is subject to the individual annotator's discretion. In general, breakpoints tend to occur every three to eight seconds.

Some things to consider when inserting timestamps of any kind:
- Timestamps must never occur in the middle of a word.
- Be careful not to clip off the end/beginning of a word when inserting a timestamp. This is trickiest with certain sounds, like "s", "f", "t", "k", "p". Take special care when inserting timestamps around words that begin or end with these sounds.

Good places to insert timestamps are
- at pauses
- at breaths
- at ends of sentences or phrases

3 Transcription

3.1 Overview

Once a file has been fully segmented, it must be transcribed. Annotators produce a verbatim (word-for-word) transcript of everything the target speaker said within each file. The words transcribed within each segment boundary must correspond exactly to the timestamps that have been created, so that the audio file is correctly aligned with the transcript. Special symbols and conventions are used to identify particular features of the transcript. These are summarized in Appendix A.

3.2 Process

Again using Transcriber and working with the segmentation file for a single speaker, annotators play each segment (each approximately 3-8 seconds long) and type what they hear. Like segmentation, transcription proceeds in two or more passes. The first pass aims to capture what the target speaker said, while the second and any consecutive passes refine the original transcript, add detail and correct errors.

3.3 Transcription Conventions

3.3.1 Orthography and spelling

3.3.1.1 Capitalization

Capitalization in the transcripts is used to aid human comprehension of the text. Annotators should follow standard written capitalization patterns, and capitalize words at the beginning of a sentence, proper names, and so on.

3.3.1.2 Spelling

Transcribers use standard orthography, word segmentation and word spelling. All files are spell-checked 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 or other reference material.

3.3.1.3 Contractions

Annotators limit their use of contractions to those that exist in standard written English, and of course only when a contraction is actually produced by the speaker. Annotators should take care to transcribe exactly what the speaker says, not what they expect to hear. 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**	should of, shoulda
Would have	***would've***	**would've**	would of, woulda
it is	***it's***	**it's**	its
its (possessive)	***its***	**its**	it's
Marvin (possessive)	***Marvin's***	**Marvin's**	
Marvin is	***Marvin's***	**Marvin's**	
Marvin has	***Marvin's***	**Marvin's**	
Going to	***gonna***	**going to**	gonna
want to	***wanna***	**want to**	wanna
got to	***gotta***	**got to**	gotta

Note: Annotators should take care to avoid the common mistakes of transposing possessive its for contraction it's (it is); possessive your for the contraction you're (you are); and their (possessive), they're (they are) and there.

Annotators should transcribe exactly what they hear using standard orthography. If a speaker uses a contraction, the word is transcribed as contracted: they're, won't, isn't, don't and so on. If the speaker uses a complete form, the annotator should transcribe what is heard: they are, is not and so on.

For non-standard contractions like "gonna" and "wanna" annotators should spell out the entire word: going to, want to.

3.3.1.4 Numbers

All numerals are written out as complete words. Hyphenation is used for numbers between twenty-one and ninety-nine only.

```
twenty-two
nineteen ninety-five
seven thousand two hundred seventy-five
nineteen oh nine
```

3.3.1.5 Hyphenated words and compounds

In general, annotators should be conservative about use of hyphens. For instance:

an overly complicated analysis **not** *an overly-complicated analysis*

However, in some cases, a hyphen is required:

anti-nuclear protests **not** *anti nuclear protests*

Compounds can be tricky. When in doubt, annotators should consult a dictionary.

3.3.1.6 Abbreviations

In general abbreviations should be avoided and words should be transcribed exactly as spoken. The exception is that when abbreviations are used as part of a personal title, they remain as abbreviations, as in standard writing:

```
Mr. Brown
Mrs. Jones
Dr. Spock
```

However, when they are used in any other context, they are written out in full:

```
I went to the junior league game.
I went to the doctor, and all he said was, don't worry, it's
natural.
Hey mister, do you know how to get to the stadium?
```

3.3.1.7 Acronyms and spoken letters

Acronyms that are normally written as a single word but pronounced as a sequence of individual letters should be written in all caps, with each individual letter preceded by a ~ tilde symbol:

```
~S ~P ~C ~A
~C ~E ~O
```

Similarly, individual letters that are pronounced as such should be written in caps, with each letter preceded by a tilde:

```
I got an ~A on the test.
His name is spelled ~S ~I ~M ~P ~S ~O ~N.
```

3.3.1.8 Punctuation

Annotators should use standard punctuation for ease of transcription and reading comprehension. Acceptable punctuation is limited to periods and question marks at the end of a sentence, and commas within a sentence. Transcripts should not contain quotation marks, exclamation marks, colons, semicolons, dashes or ellipses. Punctuation is written as it normally appears in standard writing, with no additional spaces around the punctuation marks.

3.3.2 Disfluent speech

3.3.2.1 Introduction

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 numerous of hesitation sounds. Annotators should take particular care in sections of disfluent speech to transcribe exactly what is spoken, including all of the partial words, repetitions and filled pauses used by the speaker.

3.3.2.2 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. Each language has a limited set of filled pauses that speakers can employ. 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"), but rather is restricted to these five non-lexemes:

```
ah              uh
eh              um
er
```

3.3.2.3 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.

```
Yes, absolu- absolutely.
```

3.3.2.4 Restarts

Speaker restarts are indicated with double dash --. Annotators use this convention for cases where a speaker stops short, cutting him/herself off before continuing with or rephrasing the utterance.

```
I can't really say that there should be a -- what type punishment
there should be.
```

3.3.2.5 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.

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

3.3.3 Additional markup

3.3.3.1 Hard-to-understand sections

Sometimes an audio file will contain a section of speech that is difficult or impossible to understand. In these cases, annotators use double parentheses (( )) to mark the region of difficulty.

Sometimes it is possible to take a guess about the speaker's words. In these cases, annotators transcribe what they think they hear and surround the stretch of uncertain transcription with double parentheses:

```
And she told me that ((I should just leave.))
```

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. Where possible, this untranscribed region gets its own timestamp.

3.3.3.2 Non-standard words

Occasionally a speaker will make up a new word on the spot, or will use a non-standard idiosyncratic or regional word that is not part of the general language. These are not the same as slang words or mispronunciations; they are words that are unique to the speaker in that conversation or are unique to the dialect. If annotators encounter such a non-standard 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?
They have as much *knowledgement about things as we've got.
```

3.3.3.3 Foreign languages

Portions of speech in another language are annotated using the <language text> convention to indicate the language and to transcribe the words that are spoken in that language. For instance:

```
And then I took all of the <German Sachen> to my room.
Oh, <Spanish gracias> he said.
```

3.3.3.4 Interjections

The following standardized spellings are used to transcribe interjections. Interjections do not require any special symbol.

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

3.4 Some general considerations

Annotators should not try to correct non-standard grammatical features; e.g. "I seen him" for "I saw him" should be transcribed as spoken. The same goes for words that are used in a non-standard way: annotators should transcribe what is spoken, not what they expect to hear.

Annotators should not try to imitate a speaker's non-standard pronunciation. Standard spelling should be adopted for non-standard pronunciations. Obviously mispronounced words (as opposed to non-standard pronunciations) should be marked with the plus + symbol.

4 Second Passing

4.1 Introduction

Second passing is used as a quality control measure to ensure the accuracy of segmentation, transcription including markup, and speaker identification. After the initial file has been fully segmented and transcribed, a new annotator listens to the entire interview while viewing the corresponding transcript and makes adjustments to the timestamps or transcription as needed. Second passing entails a combination of manual and programmatic checks on the transcript files. The particular types of checks conducted during second passing are described below.

4.2 Segmentation

Second pass annotators verify that each timestamp matches the corresponding transcript exactly. Annotators play each timestamp in turn and confirm that the audio and transcript for that segment are an exact match and make any necessary corrections. Annotators also check that the timestamp has been placed in a suitable location, e.g. between phrases or sentences, and that the timestamp does not chop off the start or end of any word.

Annotators listen to the entire file to ensure that all speech for each file is captured within a turn segment and that no speech remains outside of a segment boundary.

4.3 Transcription verification

During the transcript-checking phase of second passing, annotators examine the transcript in detail, checking for accuracy, completeness and the consistent use of transcription conventions. Annotators pay close attention to a handful of areas that are especially difficult to transcribe, in particular unintelligible speech sections and areas of speaker disfluency. Any proper names whose spelling could not be verified during the initial transcription process are corrected and standardized within the file. Finally, annotators conduct a spell check on the file.

4.3.1 Dialect-specific transcript verification

Because of the nature of the SLX data, one additional pass was conducted over a portion of the data. For those interviews where the main subject is a speaker of British rather than American English, a native British English speaking sociolinguist reviewed the transcripts once again to check for errors caused by the original transcribers' (all American English speakers) unfamiliarity with cultural context, slang or pronunciation.

In the examples below, the first column presents the original transcriber's transcription; the second column demonstrates the native British English speaker's revision of unintelligible speech portions:

`Is that ((Hugh Potty))?`	`Is that how you put it?`
`She done her lovely.`	`She done a wobbler.`
`Bloody (( )) uh.`	`Bloody nutters, youse are.`
`All ((amber)) heads.`	`All them birds.`

4.3.2 Treatment of personal names and identifying information

As a final stage in the processing of transcript files, annotators specially label all personal names and other personal identifying information contained in each interview. This includes full names of the interview participants, street addresses or other information that might reveal the true identity of the speakers in the interviews. First names in isolation or in narratives and mentions of places or people that are not revealing of a speaker's identity (or the identity of the speaker's personal contacts) are not labeled. All personal names tagged in this way are then replaced with a standardized text string #Name_suppressed# and separate timestamps are created to surround the name. Finally, the corresponding segments in the audio files are modified to "bleep out" the personal names. The bleeping process preserves some of the prosodic structure of the speech while obscuring the phonemic content.

Appendix A: Summary of special symbols

Category	Condition	Markup	Example	Explanation
Orthography and spelling	**Numbers**	Spelled out	twenty-five, one oh nine, one hundred thirty-seven	Write out in full; dashes 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	, ? .	Limited to these three symbols.
	Pronounced acronyms	@	@NAFTA	Write letters with all caps, no space between letters.
	Individual letters	~	~I before ~E ~Y ~M ~C ~A	Individual letters spelled out, capitalized, each with ~
Disfluent speech	**Filled pauses**	No special markup	ah, eh, er, oh, uh	Limited to this list.
	Partial words	-	absolu-	Speaker-produced partial words are indicated with a dash. Transcribe as much of the word as you hear.
	Speaker restart	--	I thought he -- I thought he was there.	Used when the speaker stops short and then repeats him/herself, or abandons the utterance completely, restarting with a new sentence.
	Mispronounced words	+	+probably	Mispronounced word (a speech error). **NOTE**: 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.
Other markup	**Unintelligible speech**	(( ))	(( ))	This indicates an entirely unintelligible passage.
	Semi-intelligible speech	((text))	They lived ((next door to us)).	This is the transcriber's best attempt at transcribing a difficult passage.
	Non-standard words	*	*poodleish	Speaker uses an idiosyncratic or dialect-specific word.
	Foreign language	<language text>	<French merci>	This is used to indicate foreign speech. If the word is unknown, leave it out. **NOTE**: Do not use this convention for foreign borrowings that are common in the target language, e.g. *apropos*.
	Interjections	no special markup	uh-huh, yeah, mhm	Use standardized spellings; limited to a short list.