

Spontaneous Speech Transcription Guidelines for Speech Accessibility Project

May, 2023: Phase 1, Parkinson's Disease

Introduction

- The transcription is verbatim. All words you hear, including false starts, filler words, and repetitions in normal disfluencies, should be typed as they are. The exceptions are stuttering-repetitions, which will be transcribed using intelligent verbatim. Details are below.
- Standard American written spelling, capitalization, and punctuation should be used.
- No segmentation will be done.

Capitalization

- The beginning of sentences and proper names should be capitalized.
 - Examples
 - *I've been reading a book about Sidney Poitier, a great actor.*
 - *Morning is coffee*
- Spoken letters and acronyms should be capitalized: more details below.

Individual Spoken Letters and Acronyms

- Individual spoken letters should be transcribed in caps and preceded by a tilde symbol '~'.
 - Examples
 - *One of my favorite pets is named Jessye. We spelled it ~J ~E ~S ~S ~Y ~E, after the opera singer.*
 - *A pangolin is my favorite animal. ~P ~A ~N ~G ~O ~L ~I ~N*
- Acronyms that are pronounced as a series of letters should be preceded by '~'. However, well-known named entities or trademarks can be written in their Wikipedia standard spelling without '~'. When uncertain, check to see if the entity exists on Wikipedia.
 - Examples
 - *My favorite sandwich is a ~BLT.*
 - *specifically ~WHYY, Philadelphia's ~NPR station*
 - *U2 (in stead of '~U two')*
- Acronyms that are pronounced as words should be written as such, with no special markup.
 - Examples
 - *ASAP*
 - *HIPAA*

Numbers

- All numerals must be written out as complete words according to how they are spoken.
- Hyphenate numbers between twenty-one and ninety-nine.
 - Examples
 - *My favorite television show was The Dark Shadows back in the sixties.*
 - *Put all of his money on red twenty-three*

Contractions

- Transcribe contractions such as 'I'm', 'You're', and 'gonna' according to standard spelling.
 - Examples
 - *I'm staying up later at night.*
 - *It's actually changing*
 - *You're in a safe place*
 - *This is gonna be difficult.*

Abbreviations

- Abbreviations should be avoided and words should be transcribed exactly as spoken. However, when abbreviations are used as part of a personal title, they should be transcribed as abbreviations.
 - Examples
 - *Dr. Michael Olken is one of the authors of it.*
 - *I just saw the film again Mr. and Mrs. Smith*

Disfluencies

Disfluencies are all marked using parentheses.

- **False starts** (sentences and words that are cut short before completion): They should be transcribed as spoken. They should not be omitted. Mark them using parentheses.
 - Examples
 - *(It's in the) It takes place in the future*
 - *(I mean there's) Ask him some questions about that.*
 - *but (um I just get) the time gets away from me*
- **Repetitions of full-words or phrases:** They should be transcribed and marked using parentheses.
 - Examples
 - *(And) And we would go on lots of trips*
 - *what Boston looked like (back) back then*
 - *(What's the what's the) What's the weather like today?*
- **Partial words, either a single sound or syllables, due to repetitions or repairs:** When a speaker breaks off in the middle of the word, whether that results from repetitions or repairs, partial words should be transcribed using parentheses and a single hyphen '-', without spacing between the hyphen and the word. The hyphen indicates the point at which the word was broken off.
 - Examples
 - *My favorite (s-) sandwich is made by toasting bread*
 - *I'll choose the best of the (a- a-) available choices*
 - *canned (tur-) turkey*
 - *I watched every one of the (reserv-) renovation*
 - *take my (frien-) family with me on these*
 - *portions of the (flow-) plant*

- When a single sound is repeated, a letter that represents the sound should be transcribed, instead of the first letter of the spelling, because the first letter can have more than one pronunciation.
 - Examples
 - *from a (k-) concert of the Kingston Trio.*
 - (k-), instead of (c-), is used even though it's a partial word of 'concert' since the letter 'c' is ambiguous with two possible pronunciation either the [k] sound (as in 'cat') or the [s] sound (as in 'city').
 - *I like (sh- sh-) Chicago.*
 - *she'll go the (h-) whole way with me*
 - Sound prolongation is not annotated.
- **Stuttering-repetitions:** Unlike repetitions in normal disfluencies, we use intelligent verbatim, i.e., the number of times it's repeated is not transcribed. Stuttering-repetitions are marked using parentheses and a single hyphen '-', followed by an asterisk '*', without spacing between the hyphen and the asterisk symbol.
 - Examples
 - *(f-*) Find m-my phone.*
 - *What's the (w-*) weather like today?*
- **Filler words:** Speakers may use filler words such as 'um, uh, ah' to indicate hesitation or to indicate they are thinking of what to say next. All filler words should be transcribed, using the standardized spellings. Mark them using parentheses.
 - Examples
 - *And (uh) it's a dry humor.*
 - *I value kindness, (uh) good communication skills*

Interjections

- Interjections that convey reactions or sudden expressions usually occur in conversation, so we expect these to occur rarely. However, in case they appear, they should be transcribed, using the standardized spellings: e.g., jeez; huh; oh; okay; uh-huh; uh-oh; whoa; whew; yeah. They are NOT marked with parentheses.
 - Examples
 - *yeah, I think that pretty much says it all*
 - *Hm that's a funny question.*
 - *whoa Linda Ronstadt will be our first lady!*

Personally Identifiable Information (PII): Privacy

- If speakers say their name (the first name, or the last name, or the full name), none should be transcribed to protect the privacy of the speaker. We consider the first name alone as a potential PII since disease diagnosis is an example of PII that may in combination allow a person to be identified and our participants frequently mention their diagnosis and city names, which we do not remove as PII.
- If speakers say the name of family members, none should be transcribed either.

- If speakers say their addresses, ZIP codes, phone numbers, and email addresses, none should be transcribed.
- The tag [PII] is used to indicate the space of the PII. In the corresponding audio file, the section should be replaced with a silent segment.
 - Examples
 - *"[PII], we'd like to hear from you".*
 - *And I feed [PII] and get her her meds.*
 - *together with [PII], our son*

Hard-to-understand sections: words or phrases you don't understand

If there is a section of speech that is hard to understand, the section should be marked using curly brackets { }. Depending on the degree of uncertainty, there are three different codes.

- Your best guess about the word spoken should be marked using {g: WORD}, that is, curly brackets, the letter 'g', a colon, a space, and then the guessed word(s).
 - Examples
 - *you {g: loosen} down, {g: pay} more and then off you go*
 - *And she was my {g: idol}.*
- When it's impossible to make a guess what words were spoken, your best guess about how many words were spoken should be marked using {w: NUMBER}, that is, curly brackets, the letter 'w', a colon, a space, and the guessed number.
 - Examples
 - *There was a short article and everything was {w: 3}*
 - *I like to travel by {w: 3} {instead of} by plane*
- When it's impossible to make a guess even about how many words were spoken, mark it using {u: }, that is, curly brackets, the letter 'u', a colon, and a space.
 - Examples
 - *Replace {u: } very tightly as tight as possible.*
- Note. In the earlier version of transcription guidelines (March 2023), hard-to-understand sections were simply marked using { }, without any characters such as 'g:' 'w:' and 'u:'. The earlier transcripts will be gradually updated with the conventions in this version.

Punctuation

- Use periods, question marks, exclamation marks, commas, and quotation marks following the conventional standards.
- The end of utterances is marked by periods, question marks, or exclamation marks.
 - Examples
 - *He said, "Are you working, or are you fishing?"*
 - *Mix that up in there as well and voila!*
 - *Do I like to travel? Of course!*
- Commas are used for listing and especially when it would be confusing to read this part of the transcript without a comma.
 - Examples
 - *Then, you put mustard, then ham, then salami.*

Errors and Non-Standard Usage (COLLOQUIALISMS)

- Do not paraphrase, nor correct grammatical errors. Transcribe what is spoken.

We do not mark pauses.

We do not mark background noises in the transcript.

- We note any notable noise in the Comment section.

We do not mark non-speech speaker noises such as laughter in the transcript.

- This will change in Phase 2. The two non-speech speaker noises will be transcribed: [cough] and [throat clearing].

Prompts for spontaneous speech

In the transcript, spontaneous speech prompts are not deleted, but marked using square brackets [] for reference purposes.

- Examples
 - *[Talk about a news story you remember from childhood.] A news story that I remember from my childhood happened in (uh), I think, November nineteen sixty-three.*

Additional guidelines about reading prompts, invalid audio files, and SLP comments.

- Reading prompts

In addition to spontaneous speech, there are two types of reading prompts (digital assistant commands and novel sentences). The main transcription tasks for reading prompts are to verify that speakers have correctly read the text prompt, and to edit the transcript to reflect errors (e.g., if a word is omitted) so that the transcript matches what speakers actually said.

Disfluencies should be marked the same way as spontaneous speech.

- Invalid audio files

When audio files are not valid, the following codes are used to indicate the reason for exclusion of the files in the database.

- **nod**: no data.
 - This is when there is no sound and a .wav file does not even exist.
- **nov**: no voice.
 - It may have some noise or not.
- **bgn**: background noise.
 - This is when voice is present but loud background noise completely mask the voice.
- **otv**: presence of other voice.
 - This is when there is a voice, or even multiple voices, but they are not from the participant.
- **cut**: voice cut off.
 - This is when the voice is less than one word.

- SLP assessment comments

When SLP comments are added in the comment fields, they are marked using square brackets.

- Examples
 - [Pace more rapid than would be expected with a digital assistant.]
 - [Creaky voice.]

Summary of the guidelines with examples

Category	Instruction	Examples
Capitalization	The beginning of sentences, proper names, individual spoken letters and acronyms are capitalized	<ul style="list-style-type: none"> • <i>Sidney Poitier</i> • <i>Morning is coffee.</i>
Individual spoken letters	Capitalized and preceded by a tilde symbol ~	<ul style="list-style-type: none"> • ~J ~E ~S ~S ~Y ~E • ~P ~A ~N ~G ~O ~L ~I ~N
Acronyms pronounced as a series of letters	Capitalized and preceded by a tilde symbol ~	<ul style="list-style-type: none"> • ~WHYY, Philadelphia's ~NPR station
Acronyms pronounced as words	Written as such, with no special markup	<ul style="list-style-type: none"> • ASAP • HIPAA
Numbers	Written out. Numbers between twenty-one and ninety-nine are hyphenated	<ul style="list-style-type: none"> • <i>in the sixties</i> • <i>twenty-three sixties</i>
Contractions	Standard spelling	<ul style="list-style-type: none"> • <i>I'm staying up later</i> • <i>This is gonna be difficult</i>
Abbreviations	Personal titles are transcribed as abbreviations	<ul style="list-style-type: none"> • <i>Dr. Michael Olken</i> • <i>Mr. and Mrs. Smith</i>
Disfluencies	False starts are transcribed as spoken, with no special codes except parentheses	<ul style="list-style-type: none"> • <i>(They could) if they just waited another minute</i> • <i>(I mean there's) Ask him some questions about that.</i>
	Repetitions of full-words or phrases are transcribed as spoken, with no special codes except parentheses	<ul style="list-style-type: none"> • <i>(back) back then</i> • <i>(What's the what's the) What's the weather like today?</i>
	Partial words are transcribed as spoken, with a single hyphen and parentheses	<ul style="list-style-type: none"> • <i>My favorite (s-) sandwich</i> • <i>canned (tur-) turkey</i> • <i>I watched every one of the (reserv-) renovation</i>
	Stuttering-repetitions are transcribed using parentheses and a single hyphen, followed by an asterisk *	<ul style="list-style-type: none"> • <i>(f-*) Find m-my phone.</i> • <i>What's the (w-*) weather like today?</i>
	Filler words (hesitation sounds) are transcribed using standard spellings and parentheses	<ul style="list-style-type: none"> • <i>(uh) it's a dry humor.</i> • <i>I value kindness, (uh) good communication skills</i>
Interjections	Transcribed using standard spellings. No special marking	<ul style="list-style-type: none"> • <i>yeah, I think that pretty much says it all</i> • <i>whoa Linda Ronstadt will be our first lady!</i>
Personally Identifiable Information (PII)	<p>The following PII are marked as [PII]:</p> <ul style="list-style-type: none"> • A speaker's first name, last name, full name, and the name of family members. • A speaker's addresses, ZIP codes, phone numbers, and email addresses. <p>In the corresponding audio file, the PII section is replaced with a silent segment.</p>	<ul style="list-style-type: none"> • <i>"[PII], we'd like to hear from you".</i> • <i>together with [PII], our son,</i>

Hard-to-understand sections: words or phrases you don't understand	Best guess about the word spoken is marked using {g: WORD}		<ul style="list-style-type: none"> • <i>she was my {g: idol}.</i>
	When it's impossible to guess what words were spoken, best guess about how many words were spoken is marked using {w: NUMBER}		<ul style="list-style-type: none"> • <i>There was a short article and everything was {w: 3}</i>
	When it is impossible to make a guess even about how many words were spoken at what was said, the section is marked using {u: }		<ul style="list-style-type: none"> • <i>Replace {u: } very tightly as tight as possible.</i>
Punctuation	The end of utterances is marked by periods, question marks, exclamation marks. Commas are used for listing and especially when it would be confusing to read this part of the transcript without a comma		<ul style="list-style-type: none"> • <i>He said, "Are you working, or are you fishing?"</i> • <i>Of course!</i> • <i>Then, you put mustard, then ham, then salami.</i>
Summary of non-letter characters to code specific meanings	~	individually spoken letters	See above examples
	()	all types of disfluency	
	-	partial words when used inside parentheses	
	*	stuttering-repetitions	
	{ }	hard-to-understand sections	
	[]	spontaneous speech prompts, PII, and SLP comments, i.e., no associated audio exists.	