

# Speech Transcription Guidelines for Speech Accessibility Project

## Phase 2

Target etiologies: Parkinson's, amyotrophic lateral sclerosis, cerebral palsy, Down syndrome, stroke

### Introduction

- The transcription is verbatim. Do not paraphrase, nor correct grammatical errors. 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.

### 1. Capitalization

- 1.1. The beginning of sentences and proper names should be capitalized.
  - e.g., *I've been reading a book about Sidney Poitier, a great actor.*
  - e.g., *Morning is coffee.*
- 1.2. Spoken letters and acronyms should be capitalized: more details in section 3.

### 2. Punctuation

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

### 3. Individual Spoken Letters and Acronyms

- 3.1. Individual spoken letters should be transcribed in caps and preceded by a tilde symbol '~'.
  - e.g., *One of my favorite pets is named Jessye. We spelled it ~J ~E ~S ~S ~Y ~E, after the opera singer.*
  - e.g., *A pangolin is my favorite animal. ~P ~A ~N ~G ~O ~L ~I ~N.*
- 3.2. 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.

- e.g., *specifically ~WHYY, Philadelphia's NPR station*
- e.g., *Recently I saw the movie 'Go On, Be Brave' about a woman with ALS.*

3.3. Acronyms that are pronounced as words should be written as such, with no special markup.

- e.g., *HIPAA*

3.4. Exceptions: There are situations where acronyms that are commonly pronounced as a series of letters are actually spoken as words. These are new cases involving non-native speakers from Puerto Rico and other etiologies. They are transcribed with only the first letter of the word in uppercase.

- e.g., *Ada compliant*

#### 4. Numbers

4.1. All numerals must be written out as complete words according to how they are spoken.

4.2. Hyphenate numbers between twenty-one and ninety-nine.

- e.g., *My favorite television show was The Dark Shadows back in the sixties.*
- e.g., *Put all of his money on red twenty-three.*

#### 5. Contractions

5.1. Transcribe contractions such as 'I'm', 'You're', and 'gonna' according to standard spelling.

- e.g., *I'm staying up later at night.*
- e.g., *You're in a safe place*
- e.g., *This is gonna be difficult.*

#### 6. Abbreviations

6.1. 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.

- e.g., *Dr. Michael Olken is one of the authors of it.*
- e.g., *I just saw the film again Mr. and Mrs. Smith*

#### 7. Disfluencies

7.1. Disfluencies are all marked using parentheses.

7.2. **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.

- e.g., *(It's in the) It takes place in the future*
- e.g., *(I mean there's) Ask him some questions about that.*

- e.g, *but (um I just get) the time gets away from me*
  - e.g., *The main thing is my partner. (I like to) If I can go with my wife, I like it a lot more.*
- 7.2.1. When individuals have cognitive or reading challenges (e.g., aphasia secondary to stroke), ‘false starts’ are transcribed as spoken but are not marked with parentheses.
- e.g, *If animals could talk, can where did do you like you like to say?*
  - e.g, *Live your could live ~A in world, where would live, world in the world live.*
- 7.3. **Repetitions of full-words or phrases:** They should be transcribed and marked using parentheses.
- e.g, *(And) And we would go on lots of trips*
  - e.g, *what Boston looked like (back) back then*
  - e.g, *(What's the what's the) What's the weather like today?*
- 7.4. **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.
- *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*
- 7.4.1. 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.
- e.g, *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 various pronunciation such as the [k] sound (as in ‘cat’) or the [s] sound (as in ‘city’).
  - e.g, *like (sh- sh-) Chicago*
  - e.g, *she'll go the (h-) whole way with me*
- 7.5. **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.

- e.g, (f-\*)*Find (m-) my phone.*
- e.g, *What's the (w-\*)weather like today?*

7.6. **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.

- e.g, *And (uh) it's a dry humor.*
- e.g, *I value kindness, (uh) good communication skills*

## 8. Interjections

8.1. 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.

- e.g, *yeah, I think that pretty much says it all*
- e.g, *Hm that's a funny question.*
- e.g, *whoa Linda Ronstadt will be our first lady!*

## 9. Personally Identifiable Information (PII): Privacy

- 9.1. 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.
- 9.2. If speakers say the name of family members, none should be transcribed either.
- 9.3. If speakers say their addresses, ZIP codes, phone numbers, and email addresses, none should be transcribed.
- 9.4. 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.
- e.g, "[PII], we'd like to hear from you".
  - e.g, *And I feed [PII] and get her her meds.*
  - e.g, *together with [PII], our son*

## 10. Hard-to-Understand Sections: Words or Phrases You Don't Understand

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

- 10.1.1. **{g: WORD}**: 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).
- e.g., *you {g: loosen} down, {g: pay} more and then off you go*
  - e.g., *And she was my {g: idol}.*
- 10.1.2. **{w: NUMBER}**: 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.
- e.g., *There was a short article and everything was {w: 3}*
  - e.g., *I like to travel by {w: 3} {g: instead of} by plane*
- 10.1.3. **{u: }**: 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.
- e.g., *Replace {u: } very tightly as tight as possible.*
- 10.2. 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.

## 11. Sentences Spoken in a Foreign Language

- 11.1. When sentences are spoken in a foreign language, **{f: }** is used, and no annotation is provided for those sentences.
- e.g., *What is one of your favorite animals and why? (Um) {f: }*
  - e.g., *Talk about the strangest thing you ever eaten. {f: }*
- 11.2. Note: If a single word spoken in a language other than English appears in the middle of sentences, we transcribe it; they are often borrowed from French or Spanish words, typically including food names and proper nouns.
- e.g., *One of my favorite foods is croquetas. (um) They are made of flour and milk, nutmeg, and then for example (um) Spanish jamón*
  - e.g., *I was afraid to something called lagartijo. Lagartijo is like a small animal that climbs trees and stuff like that.*

## 12. Extra Voice

- 12.1. When a caregiver assists the participant during recording, their voice may be recorded along with participants' voice.
- 12.2. For the Parkinson's category, the extra voice section is marked using (cs: ).

- e.g, (cs: *Tell you when to start.*) *My family members would be my sisters and they would be similar to me.*
- e.g, (cs: *Go ahead.*) *Playing with friends is the best thing on a summer day.*
- e.g, *And that's it.* (cs: *United States.*) *Hm?* (cs: *How about the United States?*) *(Oh) Oh. We have (y-) United States a couple of places we've been to.*

12.3. For the other etiologies, the following two methods are used:

- 12.3.1. If the extra voice consists of a simple utterance such as 'Go' or 'Good', the audio section is silenced and no annotation is made for the extra voice.
- 12.3.2. When the extra voice alternates with the participant's voice throughout the recording, we use the timestamp method, in which the start time, end time, and participant's words are transcribed per line. As such, no audio edit is done. To create time-stamped transcripts, first, the intervals with the participant's voice are manually labeled in Praat. Then, time values and words are automatically written in the format below using a Praat script. '-' is used as in 'Re-' and '-quest' when a caregiver reads only part of a word at a time, and the participant repeats it accordingly, leading to a break in the middle of the word. As shown in the example below, all transcripts using the timestamp method have a header starting with #ts.

- e.g, #ts  
3.049 3.615 Re-  
4.052 4.777 -quest  
5.561 6.102 a  
6.46 7.205 taxi  
7.693 8.456 ride.

### 13. Prompts for Spontaneous Speech

13.1. In the transcript, spontaneous speech prompts are not deleted, but are marked using square brackets [ ] for reference purposes. In other words, square brackets are for holding information, such as a) PII, as described above and b) the type of prompt the participant is responding to.

- e.g, [*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.*

13.1.1. Note: If individuals first read the question and then respond to it, the read question should also be transcribed to accurately reflect what was spoken.

- e.g, [*Tell us about a book, movie, or video you have recently enjoyed.*] *Tell us about the book, movie, or video you have recently enjoyed. I read the book about the history of the church in Thessalonica.*

#### 14. What's Not Transcribed

- 14.1. We do not mark pauses.
- 14.2. We do not mark sound prolongation.
- 14.3. We do not mark background noises (e.g., people talking, dog barking, fire truck passing by, sounds from TV).
- 14.4. We do not mark non-speech speaker noises such as laughter, groaning, or throat-clearing.

#### 15. Additional Guidelines about Reading Prompts

- 15.1. 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.
- 15.2. If individuals add spontaneous speech while reading prompts, the spontaneous speech section should be marked using (ss: ), to distinguish its speaking style from the read-speech portion in the read-speech recording.
  - e.g., *Call the nearest coffee shop. (ss: I hope it's not Starbucks).*
- 15.2.1. Note: If individuals answer the reading prompt, instead of simply reading them, we transcribe them the same way as spontaneous speech. We may write 'responded as ss' in the Comment section.
- 15.2.2. Note: If spontaneous speech prompts are only read, with no responses, we transcribe what they spoke. We may write 'only read' in the Comment section.

#### 16. Invalid Audio Files

When audio files are not valid for the following reasons, they are excluded from the database. The following codes are noted in the comments to specify the reason for exclusion.

- 16.1. **nod**: no data.
  - This is when there is no sound at all and a .wav file does not even exist.
  - This was due to a technical glitch in the very initial stage, which had been fixed in March 23, 2023, thus rarely exists afterwards.
- 16.2. **nov**: no voice or no relevant voice
  - This is when there is no participant's voice at all, or their voice is present but they were not aware of the recording status, making it an accidental recording of the conversation. It may also be completely silent.
- 16.3. **bgn**: background noise.

- This is when the participant's voice is present but background noise is loud enough to mask the voice.
- 16.4. **otv:** presence of other voice(s).
- This is when there is a voice, or multiple voices, but none of them are from the participant.
- 16.5. **cut:** voice cut off.
- This is when the participant's recording is cut off abruptly, leaving less than a full word.
  - When there is an abrupt cut off, but at least one full word is produced, we do not exclude the file. Instead, we transcribe the word and include it in the database.
- 16.6. **dis:** distorted sound
- This is when the audio is severely distorted, making the speech completely unintelligible.
  - Note. When the speech is slightly distorted (e.g. echo or reverb) but still intelligible to human ears, we do NOT exclude them. We transcribe them with a header starting with #dis.
    - e.g., #dis  
How many refill are left?
- 16.7. **loa:** low amplitude
- This is when the intensity of voice is substantially low throughout the entire recording, generally defined as when amplitude is below 5dB.
- 16.8. **nrw:** no real word
- This is when the speech production contains no real words in English, a case often found in individuals with aphasia secondary to stroke.



### 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"> <li>• <i>Sidney Poitier</i></li> <li>• <i>Morning is coffee.</i></li> </ul>
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"> <li>• <i>He said, "Are you working, or are you fishing?"</i></li> <li>• <i>Of course!</i></li> <li>• <i>Then, you put mustard, then ham, then salami.</i></li> </ul>
Individual spoken letters	Capitalized and preceded by a tilde symbol ~	<ul style="list-style-type: none"> <li>• ~J ~E ~S ~S ~Y ~E</li> <li>• ~P ~A ~N ~G ~O ~L ~I ~N</li> </ul>
Acronyms pronounced as a series of letters	Capitalized and preceded by a tilde symbol ~. For well-known entities or trademarks, no need to use ~	<ul style="list-style-type: none"> <li>• ~WHYY, <i>Philadelphia's</i></li> <li>• <i>NPR station</i></li> </ul>
Acronyms pronounced as words	Written as such, with no special markup	<ul style="list-style-type: none"> <li>• <i>HIPAA</i></li> </ul>
Numbers	Written out. Numbers between twenty-one and ninety-nine are hyphenated	<ul style="list-style-type: none"> <li>• <i>in the sixties</i></li> <li>• <i>twenty-three sixties</i></li> </ul>
Contractions	Standard spelling	<ul style="list-style-type: none"> <li>• <i>I'm staying up later</i></li> <li>• <i>This is gonna be difficult</i></li> </ul>
Abbreviations	Personal titles are transcribed as abbreviations	<ul style="list-style-type: none"> <li>• <i>Dr. Michael Olken</i></li> <li>• <i>Mr. and Mrs. Smith</i></li> </ul>
Disfluencies	False starts are transcribed as spoken, with no special codes except parentheses	<ul style="list-style-type: none"> <li>• <i>(They could) if they just waited another minute</i></li> <li>• <i>(I mean there's) Ask him some questions about that.</i></li> </ul>
	Repetitions of full-words or phrases are transcribed as spoken, with no special codes except parentheses	<ul style="list-style-type: none"> <li>• <i>(back) back then</i></li> </ul>

Category	Instruction	Examples
		<ul style="list-style-type: none"> <li>• <i>(What's the what's the) What's the weather like today?</i></li> </ul>
	Partial words are transcribed as spoken, with a single hyphen and parentheses	<ul style="list-style-type: none"> <li>• <i>My favorite (s-) sandwich</i></li> <li>• <i>canned (tur-) turkey</i></li> <li>• <i>I watched every one of the (reserv-) renovation</i></li> </ul>
	Stuttering-repetitions are transcribed using parentheses and a single hyphen, followed by an asterisk *	<ul style="list-style-type: none"> <li>• <i>(f-*) Find m-my phone.</i></li> <li>• <i>What's the (w-*) weather like today?</i></li> </ul>
	Filler words (hesitation sounds) are transcribed using standard spellings and parentheses	<ul style="list-style-type: none"> <li>• <i>(uh) it's a dry humor.</i></li> <li>• <i>I value kindness, (uh) good communication skills</i></li> </ul>
Interjections	Transcribed using standard spellings. No special marking	<ul style="list-style-type: none"> <li>• <i>yeah, I think that pretty much says it all</i></li> <li>• <i>whoa Linda Ronstadt will be our first lady!</i></li> </ul>
Personally Identifiable Information (PII)	<p>The following PII are marked as [PII]:</p> <ul style="list-style-type: none"> <li>• A speaker's first name, last name, full name, and the name of family members.</li> <li>• A speaker's addresses, ZIP codes, phone numbers, and email addresses.</li> </ul> <p>In the corresponding audio file, the PII section is replaced with a silent segment.</p>	<ul style="list-style-type: none"> <li>• <i>"[PII], we'd like to hear from you".</i></li> <li>• <i>together with [PII], our son,</i></li> </ul>
Hard-to-understand sections with words or phrases you don't understand: 3 different codes, depending on the degree of uncertainty	Best guess about the word spoken is marked using {g: WORD}	<ul style="list-style-type: none"> <li>• <i>she was my {g: idol}.</i></li> </ul>
	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"> <li>• <i>There was a short article and everything was {w: 3}</i></li> </ul>

Category	Instruction		Examples
	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"> <li>Replace {u: } very tightly as tight as possible.</li> </ul>
Sentence(s) spoken in a foreign language	{f: } is used and no annotation is provided for the sentence(s) spoken in a foreign language.		<ul style="list-style-type: none"> <li>What is one of your favorite animals and why? (Um) {f: }.</li> </ul>
Extra voice	For the Parkinson's category, the extra voice section is marked using (cs: ). For the other etiologies, if the extra voice consists of a simple utterance such as 'Go', the audio section is silenced and no annotation is made. If the extra voice alternates with the participant's voice throughout the recording, we use the timestamp method (section 12.3.2), and the transcript has a header starting with #ts.		<ul style="list-style-type: none"> <li>Hm? (cs: How about the United States?) (Oh) Oh. We have (y-) United States a couple of places we've been to.</li> <li>#ts 3.049 3.615 Re- 4.052 4.777 -quest 5.561 6.102 a 6.46 7.205 taxi 7.693 8.456 ride.</li> </ul>
What's not transcribed	Pauses; background noises; non-speech speaker noises; sound prolongation.		
Additional speech during reading	If individuals add spontaneous speech while reading prompts, the spontaneous speech section is marked using (ss: ).		<ul style="list-style-type: none"> <li>e.g., Call the nearest coffee shop. (ss: I hope it's not Starbucks).</li> </ul>
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: 3 types	

Category	Instruction		Examples
		{g: }, {w: }, {u: }	
	{f: }	Sentences spoken in a foreign language	
	(cs: )	Caregiver speech	
	(ss: )	Spontaneous speech during a recording of read speech	
	[ ]	spontaneous speech prompts and PII, i.e., no associated audio exists.	

**Appendix: Revision history**

Date	Edited by	Notes
May, 2023	Heejin Kim	Initial version for Parkinson's
August, 2024	Heejin Kim	Second version for 5 etiologies