

Supplemental Material S1. Task instructions for the target identification ground truth task, which were provided to the student research assistants.

Paraphasia Target Prediction in Discourse

Task Instructions

Hello and thank you for working on this task! You will be making educated guesses about what a person with aphasia (PWA) intended to say if and when they produce paraphasias in discourse, in this case, the retelling of the Cinderella story.

What is a paraphasia?

An error on a content word (i.e., noun, verb, adjective, adverb). Paraphasias can be real words or nonwords. Errors in morphological marking of a content word (e.g., plurality, tense) **do not** count as paraphasias and should be considered syntax errors.

What is a target?

A target is the word the PWA most likely intended to say when they made a word production error.

Task Overview

We have already imported from AphasiaBank transcriptions of the Cinderella story retelling task. As part of this task, you will:

1. Complete a training on word production disorders and familiarize yourself with the Cinderella storybook and expected retellings (from control subjects).
2. Watch the audio-video recording of the retelling task and predict the targets of paraphasias in context when they occur.
3. Rate how confident you are in each of your target predictions.

Training

I added a new folder to your drive called "Target Prediction." In there is a subfolder called "Training Materials" and in there are four numbered documents. Please review those documents in the order they are numbered. First read the book chapter on word production disorders (Martin, 2017), then read the story retelling protocol (available from AphasiaBank here:

<https://aphasia.talkbank.org/protocol/english/materials-aphasia/instructions.pdf>), next read what we'll be referencing as the "target" or expected retelling (Appendix 2 from Richardson & Dalton, 2016) and finally review the wordless picture book (available from AphasiaBank here:

<https://aphasia.talkbank.org/protocol/english/pictures/>) alongside the retelling.

A Little Background on the Discourse Data

The 362 Cinderella story retelling sessions in your folder represent all of the samples in AphasiaBank that contain at least one word-level error of interest to our research team. AphasiaBank has its own protocol for annotating various word-level errors, but our team is only interested in predicting targets for certain types of errors (i.e., paraphasias our automatic paraphasia classification algorithm is designed to handle). We isolated only errored productions that have the potential of falling into certain error categories, and set up those productions for target prediction. Naturally, the error type cannot be determined until we have an agreed upon target word to compare against the production.

AphasiaBank Transcriptions

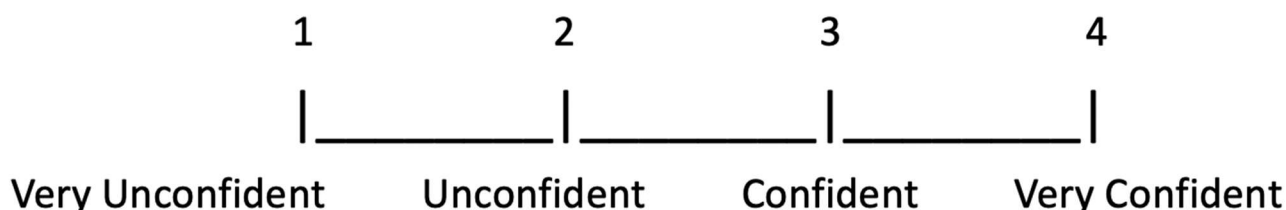
Generally speaking, all nonword paraphasias will be transcribed phonemically using IPA, and all real word paraphasias will be transcribed orthographically. AphasiaBank uses slightly different IPA conventions than our lab and/or conventions are not always consistent between participants/sessions. You should still be able to read and understand these IPA transcriptions (refer to this [clickable IPA chart](#) if in doubt). You will also have easy access to the audio recordings (time-synced at the utterance level), and should always treat the audio as the ground truth when making your predictions. There will be a chance to resolve transcriptions at a later stage in this paraphasias in discourse project.

Target Prediction

Refer to Table 9.1 in the book chapter on word production disorders to get a sense of some error patterns made by PWA. As you can see, there are numerous types of error patterns one may observe in these discourse samples. If your target prediction does not relate to the production in one of the ways outlined in the table, then please revisit your prediction process such that it follows one of these error processes. Depending on the situation, making your target prediction fit one of these processes might feel like a stretch for one reason or another. In the end, go with whatever target prediction you can make with the highest degree of confidence within this speech error framework.

Confidence Ratings

For each target prediction you make, you will attach a confidence rating (1-4) associated with that prediction. Please see below for the rating scale:



Very unconfident: no to very little support for your best guess, it is or almost is a wild guess

Unconfident: some support for your best guess, you have an argument but haven't convinced yourself it's worth defending, there's just too many unknowns and/or you have significant doubts

Confident: more support for your best guess, you have an argument and have convinced yourself it's worth defending, but you could imagine a viable counter argument

Very confident: substantial support for your best guess, you're sure or almost sure this guess is the ground truth

NOTE: Be sure to make your confidence rating based on all the information at your disposal (i.e., what you've observed within the context of the entire participant session, if applicable, and not just the amount of information available to you within the single utterance containing the paraphasia).

Target Prediction and Confidence Rating How-to's:

Using the speech error framework outlined above combined with various context clues outlined below, you will make your target predictions; and, based on the quality of clues you have to work with, assign your confidence rating. Please read through the context clues and examples below to get a better idea of how to approach this task and how to gauge your confidence level.

Context clues to help with your predictions:

- Retracings
- Syntactic and semantic clues
- Phonological similarity
- Speaker patterns
- Gestures
- Phonological fragments

Retracings

In the CHAT transcription of the utterance, retracings are often annotated or pointed out for you. Anything that has the [//] preceding it is a retracing of the original (retraced) production. For example:

Cinderella went to the **bomb** [//] ball.

Here, bomb is the errored production or paraphasia and what is retraced. Ball is the retracing and, given the context, most likely what the participant intended to say (i.e., the target). Note that retracings do not always lead to "correct" productions that make sense in context. In this particular example, though, where the retracing makes sense syntactically and semantically (it's a noun following "went to the" and a

place Cinderella goes to in the story), we would predict that the participant said bomb for the target ball and rate that prediction as "very confident."

Syntactic and Semantic Clues

Refer to the CHAT transcription and audio recording to get a sense of the context in which the paraphasia was produced. The sentence structure, if there is one, can give clues about what type of content word to expect for the target (i.e., a noun, verb, adjective, or adverb). Similarly, semantic clues in the utterance can also guide your target prediction and boost your confidence. For example:

Cinderella has a mean **dεpmam**.

The syntactic structure "has a" would lead you to predict a noun target is coming. And from what you know about the story, the adjective "mean" gives you the semantic clue that the target most likely is not one of the nice characters (e.g., prince, fairy godmother, etc.) but rather one of the mean characters (e.g., stepmother, stepsister, etc.). Since this example has strong semantic and syntactic clues and the production is phonologically similar to /stepmom/, we would rate the target prediction of /stepmom/ with the highest degree of confidence.

We would not feel as confident that the target was /stepmom/ if the production bore no or less phonological resemblance to /stepmom/. For example:

Cinderella has a mean **stεpdʒctə**.

This paraphasia may equally likely have a target of stepmother or step sister since it doesn't seem more phonologically similar to one over the other. We are still confident that it is one of the mean characters in the story that starts with /step/, and that really narrows down the possibilities of all the things Cinderella could have. Therefore, we would pick one (stepmother or step sister) and rate that prediction as confident (but not very confident).

Alternatively, if there were only syntactic clues and no recognizable semantic or phonological cues, our confidence rating would be even lower. For example:

Cinderella has a **teɪn**.

What could "tane" be? Judging from the picture book, Cinderella doesn't have a tan, that piece of information doesn't seem relevant for moving the story forward, etc. So in the absence of any other clues or leads (e.g., gestures, speaker patterns, place in the narrative, etc.), I may land on a target prediction of /tan/ and rate my confidence as "very unconfident."

Phonological similarity

In accordance with the Philadelphia Naming Test scoring protocol (Roach et al., 1996), a production is phonologically similar to the target if it shares one of the following with the target:

1. Stressed vowel (any phoneme but /ə/)
2. Final phoneme (any phoneme but /ə/)
3. First phoneme (any phoneme but /ə/)
4. 2+ phonemes in any position (all consonants + stressed vowel)
5. 1+ phoneme in the same left-to-right syllable and word position (any phoneme but /ə/)

Consonant clusters are treated as having the same word and syllable position regardless of order. Treat the vowel and r-coloring in rhotics as separate phonemes. **Do not** consider the plural morpheme (-s) or other morphological endings when judging phonological similarity.

As we have seen in the examples thus far, a production that is phonologically similar to your predicted target, may help boost your confidence in your prediction. This is especially true if the production very strongly resembles your predicted target (e.g., it's only one or two phonemes off). Alternatively, if you have an inkling for a target based on other clues, and there is some degree of phonological similarity (defined above), that similarity, even if it seems subtle, may still boost your confidence rating for that target.

As we noted previously with the *teIn* for *tan* example, phonological similarity alone might not be a strong predictor of the target. In the example below, we have stronger indicators to guide our prediction:

Cinderella lost her **cat**.

Here semantic and syntactic clues outweigh any phonological clues. From the story, we know Cinderella lost her slipper. Therefore, /slipper/ is the most likely target in this example as opposed to some other noun that may happen to be phonologically similar to the production (e.g., cactus). Cinderella did not lose her cactus, so it is very unlikely the participant intended to say so.

Speaker patterns

As you are familiarizing yourself with a given participant in the course of a session, you may pick up some speaking patterns of that individual which may help inform your target prediction process. For example, a previous or subsequent errored production from the one you're currently working on, may be more or less recognizable to a likely target, and the recognizable production may help you recognize the less recognizable one. As in the case below:

Cinderella lost her **lɪptə**.

If you encounter this error line first, it might not be immediately obvious to you that [lɪptə] is phonologically similar to /slipper/ and therefore may be a viable best guess at the target word. If you then encounter this error line:

She (.) no (.) her **slɪptə**.

You may now have pieced together enough evidence (semantic, syntactic, and phonological) to build a case for predicting /slipper/ as the target word for both paraphasias **[lɪptə]** and **[slɪptə]**. If this occurs, be sure to make your confidence rating based on all the information at your disposal, not just the amount of information available to you on the single utterance level.

Phonological fragments

Phonological fragments annotated as & in the CHAT transcript, may offer some clues as to what production the participant was going for. For example:

Cinderella lost her &s &s **lɪptə**.

The &s fragment or false start may indicate that the participant was trying to say a word with an initial s but what they actually produced may have been slightly different.

Gestures

Participants experiencing word finding difficulties will often gesture or act out what they are trying to say. This information is tremendously valuable for this task, and why it helps to play the video recording when making tough predictions.

Clue Collection Process (for tough calls)

- Play the audiovisual recording to maximize your access to context clues, especially if you have any doubts on what the target may be. Relying on the transcription alone might not be enough to get you to your best, most confident or likely prediction. For example, it will be more informative to actually hear a nonword than to read its phonemic transcription.
- Play the utterances leading up to and following the target prediction utterance you're working on. This practice can tip you off to where the participant was or is heading. Namely, it will give you the chance to look for clues that may happen to be missing from the current utterance containing the word error but still be relevant for that error's target prediction (e.g., gestures, speaking patterns, etc.). Also, you may learn where the participant is in the course of their story retelling, either more generally (the beginning, middle, end) or with enough semantic clues, maybe more specifically what particular scene or picture they are describing.

- When struggling to come up with a target prediction, consider referring to the picture book for target inspiration, and/or to whatever extent possible, try to follow along in the book with the participant's retelling.

References

- Martin, N. (2017). Disorders of word production. In I. Papathanasiou & P. Coppens (Eds.), *Aphasia and related neurogenic communication disorders* (2nd ed., pp. 169–195). Jones & Bartlett Learning.
- Richardson, J. D., & Dalton, S. G. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, 30(1), 45–73. <https://doi.org/10.1080/02687038.2015.1057891>
- Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology*, 24, 121–133.