



Word prediction in aphasia patients' procedural discourse

Karin Cox, Bruce Childers, and Rebecca Hwa

Department of Computer Science, University of Pittsburgh

Introduction

Aphasia refers to the impairments in language comprehension and/or production that may result from brain damage [1]. Technologies for persons with aphasia (PWAs) would benefit from effective methods for word prediction; however, limited computational research has been conducted in this area (e.g., see [2]).

Question: When PWAs exhibit anomia (impaired word finding), can technology help them find the right words? A comprehensive solution to this problem might leverage multiple NLP methods and input modalities (e.g., audio, video). At present, we consider the more constrained problem of identifying what language models might be most effective in predicting target words in PWAs' transcribed speech.

Planned Methods

Dataset: AphasiaBank [3] transcriptions of "procedural discourse"

- **Speakers:** PWAs (252 transcripts); Controls (180 transcripts)
- **Discourse prompt:** "Tell me how you would make a peanut butter and jelly sandwich."
- **Transcript Sample and Prediction Task:**

error annotator's correction code for error type

```
*PAR:well ‡ (.) I would got [: get] [* p:w] the bread on a plate . [+ gram]
*PAR:&um (..) get the peanu(t) butter and jelly out (re)frigerator
&=chuckles &=head:shakes and the cabinet &=laughs . [+ gram]
```

Prediction task + baseline language model:

- Training data = controls; Test data = PWAs
- Target words = suggested correction words for PWAs' errors
- Baseline language model = N-Gram
 - 4th order
 - Implemented with the OpenGrm-Ngram toolkit
 - Preceded by standard preprocessing steps:
 - tokenization → case folding → mapping of numbers to categorical labels (cardinal / ordinal / date)

Predictions for the Baseline Model

Prediction:

- Our baseline model is a simple surface analysis of the control speakers' discourse.
- We predict that its performance will be relatively weak compared to models that incorporate more sophisticated domain-specific knowledge.

Beyond the Simple Approach

Adding domain-specific knowledge:

- The control data might be augmented with sections of newswire data that are found to be most relevant for the PB&J sandwich domain (through topic modeling; see [2]).
- If this data augmentation step proves insufficient, it may be helpful to gain deeper knowledge of the task domain through a method that infers the "script" underlying the described PB&J procedure.

References

- [1] Bradley's Neurology in Clinical Practice. 7th ed. Philadelphia, PA: Elsevier. 2016.
- [2] Adams J, Bedrick S, Fergadiotis G, Gorman K, van Santen J. Target word prediction and paraphasia classification in spoken discourse. BioNLP 2017. 2017:1-8.
- [3] MacWhinney B, Fromm D, Forbes M, Holland A. AphasiaBank: Methods for studying discourse. Aphasiology. 2011 Nov 1;25(11):1286-307.