Comparing Reliability and Accuracy of Scoring Modalities for Core Lexicon Analysis Using the AphasiaBank Database

Sarah Grace Dalton[1]*, Brielle Stark[2]*, Kristen Apple[1], Davida Fromm[3], Brian MacWhinney[3], Amanda Rensch[1], & Madyson Rowedder[1]

[1]Marquette University, [2]Indiana University Bloomington, [3]Carnegie Mellon University

*sarahgrace.dalton@marquette.edu; *bcstark@iu.edu

MARQUETTE UNIVERSITY
College of Health Sciences
Speech Pathology and Audiology

Carnegie Mellon University

Introduction

- Discourse is often disrupted in individuals with aphasia (IWAs).
- Discourse is infrequently and inconsistently used clinically despite endorsement by IWAs.[1,2]
  - Most often cited barrier is time.
- More efficient ways to implement discourse analysis are needed.
- Core Lexicon (CoreLex) lists can help clinicians identify how typical the items used by clients are.
  - Checklists exist for commonly used stimuli.[3]
  - Checklists developed from large databases of controls.

Specific Aims

- To investigate the reliability of an automatic scoring procedure for core lexicon.

Methods

- A random sample of 49 transcripts from IWAs and 48 control speakers were retrieved from the AphasiaBank database.[4]
  - Tasks included 2 picture sequence stories (Broken Window and Refused Umbrella), 1 procedural task (Peanut Butter and Jelly Sandwich), 1 story retell (Cinderella), and 1 picture scene description (Cat Rescue). See **Figure 1** in middle column
  - Samples represented discourse from a range of aphasia types and severities.
- Compare two scoring modalities:
  - Gold standard – hand scoring each transcript.
  - Experimental – automated scoring using software.

Hand Scoring Rules

- Each CoreLex item is scored as 1 (present) or 0 (absent), item scores are summed to yield the CoreLex score for each task.
  - All forms of a core lexicon item should be counted (e.g., for "runs", productions of running, run, or ran would be counted).
  - Part of speech does not impact credit (e.g., "stick" produced as a noun or verb should receive credit).
  - Do NOT provide credit for synonyms (e.g., "jogs" for "runs").
  - Score any CoreLex items produced in revisions or retracing.
  - Phonological paraphasias receive credit if recognizable as the target (~50% of phonemes match).
  - Semantic paraphasias do not receive credit unless the actual production is also a CoreLex item.

Figure 1. Examples of AphasiaBank stimuli used to elicit discourse

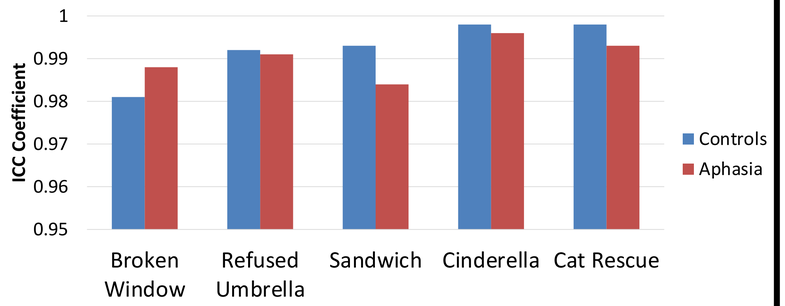
Automated Scoring

- CLAN is a free transcript analysis software (https://dali.talkbank.org/clan/)
- The CLAN command "freq +s@*filename*.cut +r6 *.cex +d2", compared transcripts for each task to CoreLex items.
  - This outputs an Excel file with participants scores for each task.

Data Analysis

- Intra-class correlation coefficients for CoreLex scores were calculated to determine absolute agreement between gold standard and experimental scoring modalities.
  - ICCs with values closer to 1 indicate better absolute agreement (e.g., hand score and automated score are very close in value).
  - ICCs were calculated separately for controls and IWAs.

Results



Discussion and Conclusions

- Both scoring methods were reliable and accurate.
- Most scoring discrepancies were resolved by editing the software command or files used to search transcripts.
- Two major sources of disagreement were identified that would be time-intensive to resolve.
  1. Information presented in revisions and retracing.
  2. Paraphasias with the intended target available.
- Automated scoring represented a significant time saver compared to hand scoring.
  - Time savings are likely to be more noticeable with increasing numbers of samples to be scored.
- Current normative data should be used with caution when comparing to automated analyses of AphasiaBank data.

References

1. Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology*, *31*(10), 1105-1126.

2. Stark, B. C., Dutta, M., Murray, L., Fromm, D., Bryant, L., Harmon, T., … Roberts, A. (2020, December 10). Spoken Discourse Assessment and Analysis in Aphasia: An International Survey of Current Practices. https://doi.org/10.31234/osf.io/v3xga

3. Dalton, S. G., Kim, H., Richardson, J. D., & Wright, H. H. (2019). A compendium of core lexicon checklists. In *Seminars in speech and language*. Thieme Medical Publishers.

4. MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, *25*(11), 1286-1307.