EasyChair Preprint

№ 6577

Determining Rater and Test-Retest Reliability of Discourse Measures in the Spoken Personal Narratives of People with Aphasia

Carla Magdalani, Brielle C. Stark, Ashleigh Doub, Anne Hittson, Julianne Alexander, Madeline Cruice, Lin Cockayne and Lucy Dipper

September 13, 2021

Determining rater and test-retest reliability of discourse measures in the spoken personal narratives of people with aphasia

Carla Magdalani[1]*, Brielle C. Stark[2]*, Ashleigh Doub[3], Anne Hittson[2], Julianne Alexander[2], Madeline Cruice[1], Lin Cockayne[1], Lucy Dipper[1]

[1]Division of Language and Communication Science, City University of London, London, UK
[2]Department of Speech, Language and Hearing Sciences, Indiana University, IN, USA
[3]Department of Speech & Hearing Science, University of Illinois at Urbana-Champaign, IL, USA

*corresponding authors, carla.chynoweth@city.ac.uk & bcstark@iu.edu

Introduction

Discourse interventions are an emerging evidence base in aphasiology [1]. However, reliably eliciting and analysing discourse is challenging for various reasons [2], and reliability is often under-reported, inadequately described, and calculated non-statistically. Notwithstanding these challenges, a stable baseline and high inter-rater reliability must be established before treatment commences to enable judgments on the effectiveness of that treatment, and to minimise errors in interpretation. In this talk, we will present evidence from two studies, which separately evaluated rater and test-retest reliability of discourse metrics in aphasia.

Methods

NEURAL Research Lab (NRL) study [3]: a virtually conducted study was conducted in 2020, recruiting 25 persons with chronic aphasia (3 excluded for significant missing data) and 24 prospectively matched adults (1 excluded for significantly poorer performance) without brain damage. Each took part in a test and a retest session, taking place 10 +/- 3 days apart, during which they told five narratives [5]. After orthographic transcription, transcripts were coded for a word-level measure, % correct information units (%CIUs).

LUNA study [4]: a virtually conducted study which began in 2020 and recruited 28 participants with chronic aphasia. Participants told and retold two personal narratives, about a week apart, prior to receiving LUNA (i.e., narrative-based) treatment. After orthographic transcription, transcripts were rated using a word-level (%narrative words) and a macrostructure-level measure (story grammar).

In both studies, two raters each analysed 50% of transcripts. 10% (LUNA) and 20% (NRL) of transcripts were randomly selected for rater reliability. Test-retest reliability was calculated for the narratives at the two time points. Intraclass correlation coefficients (ICC) analysed measures at word-level measures (excellent, >0.9; good, 0.75-0.9; moderate, 0.5-0.75; poor, <0.5) and kappa analysed the measure at macrostructure-level (very good, >0.81; good, 0.61-0.8; moderate, 0.41-0.6; fair, 0.21-0.4, poor, <0.2).

Results

Intra-rater reliability at the word-level was excellent for the LUNA study (not computed for NRL). Inter-rater reliability at the word-level was good-to-excellent for both studies. LUNA's macrostructure-level measure showed good-to-very good intra-rater reliability and moderate inter-rater reliability.

For test-retest reliability, LUNA's word-level measure showed good-to-excellent reliability, and the macrostructure-level measure showed moderate-to-good reliability. In NRL, when averaging word-level measures across all tasks, test-retest reliability was good-to-excellent for both PWA and NBD groups, though test-retest reliability ranged from poor-to-excellent when evaluated by task.

Discussion

For both studies, rater reliability was high, especially at the word-level. LUNA's macrostructure measure was analysed reliably within raters but was less reliable between raters and across timepoints. For both studies, word-level variables were found to be highly reliable in the aphasia groups. Notably, the NRL study demonstrated lower reliability for NBD group, and that reliability varied by narrative type. We will further elaborate on these results, especially clinical implications. Both studies were conducted virtually and showed high retention; this will also be discussed in the presentation.

References

[1] Dipper, L., Marshall, J., Boyle, M., Botting, N., Hersh, D., Pritchard, M., and Cruice, M. (2020). Treatment for improving discourse in aphasia: a systematic review and synthesis of the evidence base. *Aphasiology*, DOI: 10.1080/02687038.2020.1765305

[2] Cruice, M. Botting, N., Marshall, J., Boyle, M., Hersh, D., Pritchard, M., and Dipper, L (2020). UK speech and language therapists' views and reported practices of discourse analysis in aphasia rehabilitation. *International Journal of Language & Communication Disorders*, 55(3), pp. 417–442.

[3] Doub, A., Hittson, A., & Stark, B. (in press). Conducting a Virtual Study with Special Considerations for Working with Persons with Aphasia. *American Journal of Speech-Language Pathology*.

[4] Dipper, L., & Cruice, M. (2018). Personal storytelling in aphasia: a single case study of LUNA therapy. *Aphasiology*, 32(sup1), 60–61. https://doi.org/10.1080/02687038.2018.1487919

[5] MacWhinney, B., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for Studying Discourse. *Aphasiology*, 25(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893.AphasiaBank

Acknowledgments

LUNA is funded by The Stroke Association UK (TSA 2017/01). We would like to acknowledge the members of the FOQUSAphasia working group (www.foqusaphasia.com). We also acknowledge ASHFoundation New Investigator Award funding to Brielle C. Stark. This abstract is part of the Symposium entitled "Spotlighting spoken discourse in aphasia."