View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by The Aphasiology Archive

The cardinal deficit of people with aphasia (PWA) is anomia (Goodglass & Wingfield, 1997). This deficit is believed to be indicative of disruption of two cognitive processes: (i) accessing a semantic description of the target concept, and/or (ii) retrieval of a fully phonologically specified representation (e.g., Dell, 1986). During discourse, in addition to these core processes that serve word retrieval, production also depends on "…factors external to the lexicon…" (p. 169, Wilshire & McCarthy, 2002). The latter processes might influence the selection of lexical items based on syntactic, structural, and/or pragmatic criteria that can be either automatic or meta-cognitive.

One of the goals of our line of research is to investigate one aspect of how lexical items are deployed during discourse in PWA: *lexical diversity* (LD). LD is related to the range of vocabulary exhibited in a language sample (Durán, Malvern, Richards, & Chipere, 2004) and reflects a speaker's capacity to access and retrieve lexical items during discourse. One of the greatest challenges in the study of LD is the identification of a robust index to capture LD. The tools that have been frequently used by researchers are known to covary with sample length, thus yielding mathematically and conceptually spurious results (see Tweedie & Baayen, 1998).

In recent years, several novel techniques from the field of computational linguistics have been developed to assess the breadth of one's vocabulary during discourse. Though all of the techniques assert to measure LD, each one is based on its own theoretical assumptions, which is reflected in the estimation machinery they employ. Further, some of these measures have more evidence to justify the validity of their score interpretations (e.g., *D*; Durán et al., 2004; Malvern & Richards, 1997, 2000; Richards & Malvern, 1997a, 1998), some have less (e.g., Maas; Maas, 1972; *Measure of Textual Lexical Diversity [MTLD]*; McCarthy, 2005; McCarthy & Jarvis, 2010), and some have none (*Moving Average Type Token Ratio [MATTR]*; Covington & McFall, 2010) other than face validity. Also, the validity of measures that were considered the golden standard in terms of quantifying LD, such as *D* and Maas, has been questioned (Fergadiotis, 2011).

Our main goal is to supplement our understanding regarding the validity of the scores generated by different LD estimation techniques. At this time very little is known about the performance of these indices in the discourse produced by PWA. The degree to which these techniques reflect LD and little of anything else is critically related to the development of psychometrically sound measurement procedures for diagnostic and treatment efficacy purposes.

Four techniques will be explored: *D*, Maas, MTLD, and MATTR. Specific questions to be addressed include:

i. Do all the techniques generate scores that are manifestations of the same construct (i.e., LD)?
ii. Is there a single latent variable determining performance for each estimation technique or is there evidence of construct irrelevant variance?

Method

Participants. Language samples from 120 PWA from AphasiaBank, an online shared database that collects and analyzes digital recordings of the discourse of PWA across a series of tasks, are included. All participants have aphasia secondary to a left hemisphere stroke. PWA met the following inclusion criteria: (a) chronic aphasia (minimum = 6 months post onset); (b) no reported history of psychiatric or neurodegenerative disorders; (c) aided or unaided normal hearing acuity; (d) corrected or uncorrected normal visual acuity; and (e) English as their primary language. All PWA were administered the Western Aphasia Battery-Revised (Kertesz, 2007), the Boston Naming Test (Goodglass & Kaplan, 2001), and several subtests from the Reading Comprehension Battery for Aphasia, Second Edition (LaPointe & Horner, 1998).

Stimuli & Instructions. Language samples consist of responses to a story retell task designed to elicit narrative discourse (retell of the story Cinderella).

Transcription & Language Sample Preparation. Samples are digitally recorded and then orthographically transcribed in the CHAT format that is compatible with a set of programs called Computerized Language Analysis (CLAN; MacWhinney, 2000). Samples were then coded using word-level codes to indicate different types of paraphasias, repetitions, and interjections. Each word in the samples was also tagged morphosyntactically.

XML Code. Our goal is to perform a lemma-based analysis of only content words. Currently we are in the last stages of developing a set of rules using Extensible Markup Language (XML) with the following combinations of optional functions: (i) retrieve items that belong in specific word classes, (ii) retrieve lemmata or the fully inflected word forms, (iii) ignore, use, or replace paraphasias with the target words, and (iv) export output in .txt format one word per line. Once completed, the XML code would allow the user to simply define the parameters of interest to analyze multiple language samples simultaneously.

Estimating LD. Four indices of LD are selected. The first index, *D* (cf. MacWhinney, 2000) combines an algebraic transformation model and curve fitting to estimate LD and there is some evidence to support that it is relatively robust to length variation (e.g., McKee, Malvern, & Richards, 2000). The second index that is used in this study Maas (Maas, 1972) that is a logarithmic transformation of the type token ratio. Another tool that has been proposed recently for estimating LD (McCarthy, 2005) is the MTLD. MTLD reflects "…the mean length of sequential token strings in a text that maintain a given TTR value" (McCarthy & Jarvis, 2010, pp. 384). Conceptually, for any given sample, MTLD reflects how many words in a row a speaker can maintain a certain TTR. The last index of LD is the MATTR (Covington, 2010). MATTR estimates LD by using a smoothly moving window that estimates type-token ratios for each successive window of fixed length.

Preliminary Results and Discussion

Preliminary analyses have been conducted using 112 language samples. Both function words and content words were included; the analysis was not lemma-based. Two confirmatory factor analytic (CFA) models were estimated in Mplus 6.1. These included a unidimensional CFA that stipulated that every technique was a "pure" indicator of LD and a CFA model that allowed for correlated errors for the D and Maas techniques. Based on Fergadiotis (2011), the latter model assumed D and Maas scores were systematically influenced by additional factors, suggesting that they reflected something else over and above LD.

Based on several fit indices, the second model fit the data considerably better (Figures 1&2). Results indicate that even though these measures employ different computational machineries and make different theoretical assumptions, they all reflect the same construct. However, the model fit to the data adequately **only** after the error terms for D- and Maas-generated scores were allowed to covary. So, consistent with previous findings, D and Maas may reflect something else over and above the LD of the language samples (probably length effects). Importantly, the magnitude of the loadings suggests that MATTR and MTLD were the strongest indicators of the underlying trait, i.e. they reflected more strongly the variable of interest – lexical diversity.

Currently, we are in the process of finalizing the XML code and performing a lemma based analysis of content words only with the four measures. Results will be discussed with an emphasis on the clinical and research utility of the four estimation techniques.

References

Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics, 25*(2), 220-242.

Chapelle, C. A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research, 10*(2), 157-187.

Covington, M.A. (2007). *CASPR Research Report 2007-05. MATTR User Manual.*

Covington, M.A., & McFall, J.D. (2010). Cutting the Gordian Knot: The Moving- Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics, 17,* 94-100.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Goodglass, H., & Wingfield, A. (1997). *Anomia: Neuroanatomical and cognitive correlates*. San Diego, CA, US: Academic Press.

Kertesz, A. (2007). *Western aphasia battery- revised.* New York: Grune and Stratton.

Lapointe, L. L., & Horner, J. (1998). *Reading comprehension battery for aphasia.* Austin, TX: Pro-Ed.

Maas, H. D. (1972). Zusammenhang zwischen Wortschatzumfang und Länge eines Textes. *Zeitschrift für Literaturwissenschaft und Linguistik, 8,* 73-79.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk, Volume 1: Transcription format and programs* (3rd Ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

McCarthy, P. M. (2005). *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual lexical diversity* (Doctoral Dissertation). Retrieved from ProQuest Nursing & Allied Health Source database.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42(2),* 381-392.

Malvern, D. D., & Richards, B. J. (1997). A new measure of lexical diversity. In A. Ryan , & A. Wray (Eds.), *Evolving models of language* (pp. 58-71). Clevedon, UK: Multilingual Matters.

Malvern, D., Richards, B., Chipere, N., & Duran, A. (2004). *Lexical diversity and language development: Quantification and assessment* Palgrave Macmillan.

Tweedie, F. J., & Baayen, R. H. (1998). How variable may a constant be? measures of lexical richness in perspective. *Computers and the Humanities, 32*(5), 323-352.

Wilshire, C. E., & McCarthy, R. A. (2002). Evidence for a context-sensitive word retrieval disorder in a case of nonfluent aphasia. *Cognitive Neuropsychology, 19*(2), 165. doi:10.1080/02643290143000169

Figures

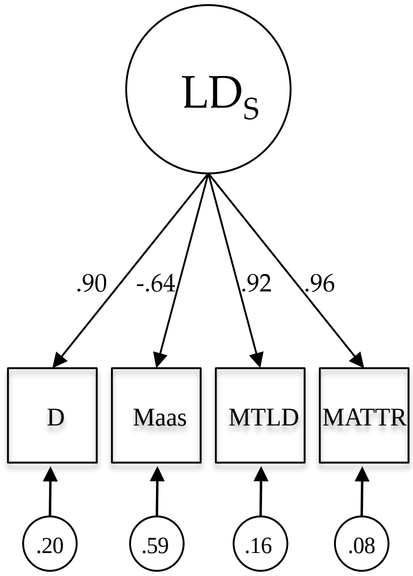

Figure 1. Model 1: χ^2 = 66.98, p < .001, Root Mean Square Error of Approximation = .54 (90% CI = .43 - .65), Standardized Root Mean Square Residual = .10

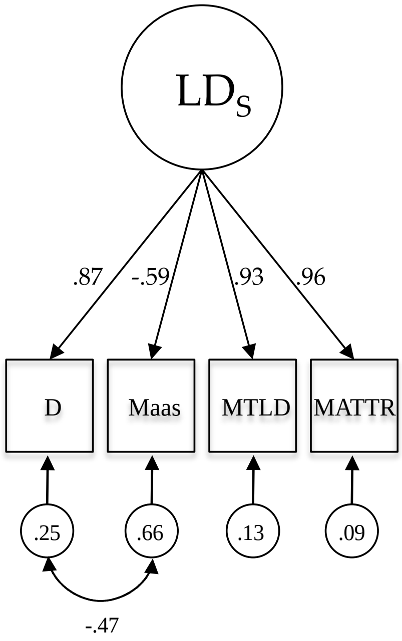

Figure 2. Model 2: χ^2 < .01, p = .93, Root Mean Square Error of Approximation < .01, (90% CI = .00 - .06), Standardized Root Mean Square Residual = .003