Core Lexicon and Main Concept Production during Picture Description

Introduction

Discourse is a highly complex and individualized communication act wherein individuals not only transmit and receive information for survival and cooperation, but also use spoken language for ritual purposes (e.g., relationships, fellowship, co-participation; Carey, 1988; Dimbleby & Burton, 1998). Discourse in typical and clinical populations has been investigated with a variety of structuralist, functionalist, and hybrid techniques (see Armstrong, 2000), and is known to be a good predictor of quality of life and life participation in persons with aphasia (PWAs). Discourse analysis, however, generally requires specialized training and can be time-consuming. MacWhinney, Fromm Holland, Forbes, & Wright (2010) suggested that analysis of a core lexicon during structured narrative tasks could provide a time-efficient and informative index of functional communication abilities. For example, clinicians could bypass lengthy transcriptions, instead generating a list of words spoken during narration for later comparison to a core lexicon (CoreLex). Using various methods, CoreLex has so far been investigated for the Cinderella story, a monologic story retell narrative task (Author1, Dillow, & Author2, 2013; MacWhinney et al., 2010) and a procedural narrative task where patients describe how to make a PB&J sandwich (Fromm, Forbes, Holland, & MacWhinney, 2013). CoreLex performance is strongly correlated with main concept (MC) production, a measure of narrative adequacy, during Cinderella retelling (Author1, Dillow, & Author2, 2013). Similar investigations for other narrative tasks are needed. The aims of this study were to 1) determine the CoreLex of a picture sequence description task included in the AphasiaBank protocol (Breaking Window), 2) calculate a CoreLex score for controls and PWAs, and 3) determine how well CoreLex predicts narrative adequacy, as judged by MC analysis.

Methods

Transcripts

Transcripts of 146 controls and 179 PWAs (56 anomic, 48 Broca's, 33 conduction, 26 NABW [not aphasic by WAB], and 15 Wernicke's) were retrieved from the AphasiaBank database. The Breaking Window picture description narrative was extracted from the transcript using the Computerized Language Analysis (CLAN) command: **gem +sWindow +n +fWindow +d1 +t*PAR +t%mor *.cha**.**

CoreLex

We first identified the entire lexicon used by controls during Breaking Window narration. The CLAN command **freq +t*PAR +s"@r-*,|-*,o-%" +o *.gem.cex +d2 -s"[+ exc]"** identified the unique lemmas in Breaking Window transcripts (N=993 lemmas). Lemmas produced by 50% or more of the control population were included in the CoreLex, with this cutoff selected because it yields a reasonably sized lexicon and it has served as criterion in previous language research (e.g., Brown's stages of development; Owens, 2008). Twenty-two lemmas were identified, and transcripts of both controls and PWAs were subsequently scored. Individuals received a "1" if the lemma was present in the transcript and a "0" if it was absent, with the sum of values across the transcript serving as the CoreLex score.

Main Concepts

Previous research identified the main concepts (MCs) produced by 50% of controls during Breaking Window narration (Author1, Campbell, Williams, Dillow & Author2, 2013). Transcripts in the current study were scored using this list of 8 Breaking Window MCs, using a coding system modified from Kong (2009) and utilized in Author1, Dillow, and Author2 (2013). A numeric value was assigned to each main concept attempt based on its accuracy and

completeness (0 = Absent, 1 = Inaccurate/Incomplete, 2= Accurate/Incomplete or Inaccurate/Complete, 3 = Accurate/Complete). These scores were then summed to yield a MC composite score for each individual.

Data Analysis

Omnibus median tests were conducted to confirm hypothesized differences between controls and PWAs for CoreLex and MC scores. Planned comparisons (median tests, Holm-Bonferroni corrected) were used to identify and characterize differences between subtypes. Spearman correlations were performed to investigate the relationship between CoreLex and MC scores.

Results

The CoreLex for Breaking Window included 22 lemmas (7 nouns, 3 verbs, 3 pronouns, 3 prepositions, 2 determinants, 1 coordinator, 1 auxiliary, 1 infinitive, and 1 copula). Omnibus median tests revealed a significant difference between control and PWAs for CoreLex, $\chi 2(5, N=325) = 195.745$, $p < .001$, and MC scores, $\chi 2(5, N=325) = 123.759$, $p<.001$. Further testing revealed significant differences between all aphasia subtypes (including those categorized as NABW) and controls for both CoreLex and MC measures ($p<.001$ for all comparisons).

Planned comparisons to examine differences between aphasia subtypes were performed following recalculation of the median (excluding controls) to avoid median inflation. Significant differences were found for both measures between Broca's aphasia and all other subtypes studied (anomic CoreLex [$\chi 2=42.298$, $p<.001$], MC [$\chi 2=26.232$, $p<.001$]; conduction CoreLex [$\chi 2=21.811$, $p<.001$], MC [$\chi 2=10.578$, $p=.001$]; NABW CoreLex [$\chi 2=61.611$, $p<.001$], MC [$\chi 2=36.626$, $p<.001$]; and Wernicke's CoreLex [$\chi 2=6.940$, $p=.008$], MC [$\chi 2=25.000$, $p<.001$]). CoreLex and MC scores also differed significantly when comparing NABW to conduction (CoreLex [$\chi 2=15.486$, $p<.001$]), MC [$\chi 2=9.502$, $p=.002$]) and Wernicke's (CoreLex [$\chi 2=8.464$, $p=.004$], MC [$\chi 2=7.031$, $p=.008$]). Finally, for CoreLex only, significant differences were observed between anomic and NABW subtypes, $\chi 2=6.451$, $p=.011$.

Using a two-tailed Spearman's correlation, CoreLex and MC scores showed a significant positive relationship after collapsing across groups, $r_s=.818$, $p<.001$. In order to identify which groups were driving this finding, we conducted the correlation for each subtype. For all subtypes except conduction ($r_s=.185$, $p<.295$), correlations were significant and positive (anomic [$r_s=.589$, $p<001$], Broca's [$r_s=.661$, $p<.001$], NABW [$r_s=.474$, $p=.014$], and Wernicke's [$r_s=.542$, $p=037$]).

Discussion

Our study provides further support that both MC and CoreLex analyses can be used for diagnosis of discourse impairments. All aphasia subtypes, including those functioning at the highest levels (i.e., NABW), differed significantly from controls. Compared to a similar investigation of the Cinderella story (Author1, Dillow, & Author2, 2013), CoreLex and MC analysis of Breaking Window narratives more consistently differentiated fluent from non-fluent aphasia types. Findings of significant differences between fluent subtypes during Breaking Window picture description were similar to that observed during Cinderella story retelling. For both Cinderella story and Breaking Window, CoreLex (but not MC) values significantly differed between anomic and NABW, indicating that though they are comparable in conveying the "gist" of the story, the typicality of the lexical items retrieved during narrative differs between the two groups.

As a picture sequence description task (4 pictures), the Breaking Window narrative is significantly shorter than the Cinderella story narrative, yet seems to have similar power in

differentiating between individuals with different subtypes of aphasia. It may be able to serve as a stand-alone discourse option in settings with time constraints (e.g., acute hospital setting) or with patients who fatigue easily. Additionally, the significant positive correlations between CoreLex and MC scores for all groups (except individuals with conduction aphasia) indicate that a CoreLex checklist may be a time-efficient and reliable predictor of narrative adequacy that is still sensitive to differences between different aphasia subtypes. However, the different correlation strengths by subtype lend support for those researchers calling for multidimensional approaches to narrative assessment. MC analysis does provide more detailed information about narrative adequacy, and the recently developed MC checklists provide a standardized, norm-referenced, and non-transcription-based method of completing such an analysis, though there are likely to be many instances when CoreLex is more practicable for clinicians to use.

References

Armstrong, A. (2000). Aphasic discourse analysis: The story so far. *Aphasiology, 14*(9), 875-892.

Carey, J. W. (1988). *Communication as culture: Essays on media and society*. New York, NY: Psychology Press.

Dimbleby, R. & Burton, G. (1998). *More than words: An introduction to communication* (3rd Ed.). New York, NY: Routledge.

Fromm, D., Forbes, M., Holland, A., & MacWhinney, B. (2013). PWAs and PBJs: Language for describing a simple procedure. Poster presented at the annual Clinical Aphasiology Conference. Tucson, Arizona.

Author1, Campbell, S. L., Williams, A. N., Dillow, E. P., & Author2. (2013). Development of clinician-friendly discourse analysis tools: Main concept analysis. Poster presented at the Annual Convention of the American Speech-Language-Hearing Association. Chicago, Illinois.

Author1, Dillow, E. P., & Author2. (2013). Core lexicon analysis: Efficient assessment of narrative discourse in persons with aphasia. Technical research session presented at Annual Convention of the American Speech-Language-Hearing Association. Chicago, Illinois.

Kong, A. P. (2009). The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: A preliminary report. *Journal of Communication Disorders, 42*(6), 442-464.

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology, 24*, 856-868.

Owens, R. E. (2008). *Language Development: An Introduction* (7th Ed.). Boston: Pearson.