12-23-2019

Moving Toward Non-transcription Based Discourse Analysis in Stable and Progressive Aphasia

Sarah Grace Dalton
Marquette University, sarahgrace.dalton@marquette.edu

H. Isabel Hubbard
University of Kentucky

Jessica D. Richardson
University of New Mexico

Seminars in Speech and Language, Vol. 41, No. 1 (2020): 32-44. DOI.

Moving Toward Non-transcription Based Discourse Analysis in Stable and Progressive Aphasia

Sarah Grace Hudspeth Dalton
Department of Speech Pathology and Audiology, Marquette University, Milwaukee, Wisconsin
H. Isabel Hubbard
Department of Communication Sciences and Disorders, University of Kentucky, Lexington, Kentucky
Jessica D. Richardson
Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, New Mexico

Abstract
Measurement of communication ability at the discourse level holds promise for predicting how well persons with stable (e.g., stroke-induced), or progressive aphasia navigate everyday communicative interactions. However, barriers to the clinical utilization of discourse measures have persisted. Recent advancements in the standardization of elicitation protocols and the existence of large databases for

development of normative references have begun to address some of these barriers. Still, time remains a consistently reported barrier by clinicians. Non-transcription based discourse measurement would reduce the time required for discourse analysis, making clinical utilization a reality. The purpose of this article is to present evidence regarding discourse measures (main concept analysis, core lexicon, and derived efficiency scores) that are well suited to non-transcription based analysis. Combined with previous research, our results suggest that these measures are sensitive to changes following stroke or neurodegenerative disease. Given the evidence, further research specifically assessing the reliability of these measures in clinical implementation is warranted.

Keywords

transcription - discourse - aphasia - clinical utility

Learning Outcomes: As a result of this activity, the reader will be able to (1) define how to score a discourse sample for main concept analysis, core lexicon, and derived efficiency; (2) describe the research evidence supporting the use of these discourse measures in persons with stable and progressive aphasia; (3) and discuss the implications of these findings on clinical utilization of discourse measures in stable and progressive aphasia.

Measurement of communication ability at the suprasentential, or discourse, level holds promise for predicting how well persons with aphasia (PWAs) navigate everyday communicative interactions. In recent years, discourse assessment has gained attention as a high-priority treatment target identified by PWAs[1] and as a primary outcome measure for aphasia treatment outcomes.[2] [3] Subsequently, investigations have sought to develop informative measures[4] [5] [6] [7] [8] and establish and/or characterize psychometric strengths and weaknesses.[9] [10] [11] Importantly, treatment studies are more frequently investigating the impact of interventions on discourse outcomes in stable, stroke-induced aphasia.[12] [13] [14] Comparatively less attention has been paid to treatment outcomes, including discourse, in primary progressive aphasia (PPA). Those studies that do examine discourse in PPA focus on word-level measures or characterization of errors[15] [16] [17] [18] [19] [20] to better understand PPA subtypes and progression, with a few exceptions.[3] [21] [22] However, PPA treatment studies are beginning to utilize discourse outcomes.[23] [24] This momentum in discourse analysis should be leveraged for a renewed focus on development or refinement of clinically relevant discourse measures. Specifically, to ensure measures can be feasibly implemented in real-world settings, barriers to clinical utilization of discourse analysis should be accounted for during development, not as an afterthought.[6] Complementary to Kagan and Simmons-Mackie's emphasis on designing treatment courses according to end-product life participation goals,[25] when designing discourse measures we should similarly progress with the end—clinical utility—in mind. We will briefly review some of the barriers to clinical utilization of discourse measurement below.

First, discourse assessment is often categorized as subjective and qualitative,[26] with historically little standardization of administration or analysis across settings. Some standardized assessments include discourse ratings, such as the Boston Diagnostic Aphasia Examination[27] or Western Aphasia Battery—Revised (WAB-R),[28] but those ratings are limited in the type of information provided (qualitative, ordinal scales, etc.) and their sensitivity to change.[29] Even when detailed instructions are included to facilitate standard administration[30] or analyses,[30] [31] [32] [33] clinicians may not be aware of their existence. If they are aware, they may have limited access or proficiency to perform computerized literature searches.[34] [35] [36] [37]

Second, limited normative data are available for interpretation of discourse samples. For example, information regarding the performance of controls on selected tasks from the stimulus set introduced by Linda Nicholas and Robert Brookshire is well documented but scattered throughout the literature for the following discourse measures: number of words and words per minute,[31] [38] correct information units (CIUs),[31] [38] [39] main concepts,[6] [32] [40] main events,[39] [41] global coherence ratings,[42] and lexical diversity.[43] There are also numerous reports of control performance for other discourse tasks and measures, for example, global coherence ratings and core lexicon for wordless picture books, personal recounts, and procedural discourse,[4] [8] cohesion measures for story retell tasks,[44] content units for picture scene description,[45] and the presence and completeness of concepts and topic coherence for picture scene description.[46] Beyond the challenge of wide dispersion, additional shortcomings include small control sample sizes, restricted age groups that do not span adulthood, and little to no evaluation of distribution properties to support their use as a normative reference. Most often, control performance has been documented primarily in evaluations of differences between age groups or differences between controls and clinical populations, and not to establish normative references or relevancy.

The aforementioned limitations are being addressed in part by the AphasiaBank database (aphasia.talkbank.org). AphasiaBank consists of hundreds of transcripts of PWAs and healthy controls contributed by researchers across the country, facilitating previously unattainable discourse analyses. Contributors utilize a standardized elicitation protocol including two free speech and five semispontaneous speech tasks, described in the studies of MacWhinney et al[47] and Forbes et al,[48] and at aphasia.talkbank.org/protocol/. The semispontaneous tasks include a picture description (Cat Rescue), two picture sequence descriptions (Broken Window and Refused Umbrella), a story retell (Cinderella), and a procedural task, “how to make a peanut butter and jelly sandwich” (Sandwich). An update of AphasiaBank resources is available in this issue.[49] Among these is a dedicated web page for publications and discourse measures using AphasiaBank data (aphasia.talkbank.org/discourse/). Through AphasiaBank, clinicians have access to standardized elicitation protocols, standardized measures with scoring information, and normative data that can aid in clinical utilization.

Finally, time is a commonly reported assessment barrier in allied health[50] (see also the article by Kim and Wright[51]). Speech-language pathologists (SLPs) report spending 1 hour or less performing outcomes assessment for each PWA on their caseload and list time constraints as the greatest barrier.[26] Discourse assessment is rarely included in that hour, with only 4% of SLPs reporting its use.[26] When considering the time required for specialized training, assessment, transcription, and interpretation, it is not surprising that discourse assessment is infrequently utilized.[52] [53] [54] In a survey of UK-based SLPs, approximately 80% of respondents reported that increased time was needed to make discourse analysis clinically feasible.[55] The most time-consuming aspect of discourse analysis is the transcription of discourse samples, which often entails phonetic transcription and error coding of paraphasias, repairs, revisions, etc. In fact, approximately 60% of respondents listed transcription as the greatest barrier to discourse analysis.[55] Furthermore, different analytic methods have different training and time requirements, which SLPs have little time for in real-world settings.

Given these barriers, to make discourse assessment more commonplace in regular practice settings, clinicians need more protected time for assessments (and associated training) and/or discourse analysis needs to take less time. Given the realities of the U.S. healthcare system, protecting clinician time, while probably the best solution, is not realistic. However, reducing the time spent on discourse assessment by decreasing transcription and/or analysis time *is* possible without enacting systems-level change. Non-transcription based discourse measures that are reliable, informative, and efficient are needed. Main concept analysis, core lexicon, and derived efficiency scores are ideal measures for ushering non-transcription based discourse analysis into the clinical setting.

Main Concept Analysis

The main concept analysis (MCA) was developed to characterize how effectively PWAs communicated the gist of a discourse task.[32] [40] Nicholas and Brookshire developed main concept checklists for a set of tasks based on clinicians' judgments of the most important story elements (i.e., story gist). Main concepts are defined as a single main verb, its subject and objects (if applicable), and any associated clauses. To complete an MCA, individual's utterances are examined for the essential elements of each main concept (i.e., verb, subject, object, clauses), and coded for presence and accuracy. Utterances that correspond to main concepts receive one of the following codes: AC—accurate/complete, all essential elements are produced and are correct; AI—accurate/incomplete, one or more essential elements is not produced, but those that are present are correct; IC—inaccurate/complete, all essential elements are produced, but one or more elements is inaccurate; II—inaccurate/incomplete, one or more essential elements is not produced, and one or more of the produced elements is inaccurate. If no utterance corresponds to a main concept, it is coded as absent (AB). The authors included detailed scoring rules in their Appendix A, which facilitates reliable coding.[32] Additionally, a formula for converting codes into scores has been developed by Kong,[56] and modified by Richardson and Dalton.[6]

MCA is sensitive to differences between clinical and control populations,[32] [56] [57] [58] [59] [60] [61] and correlates with listeners' perceptions.[62] Additionally, while different measures have been used (e.g., tau, absolute agreement), all investigations reporting reliability show adequate inter- and intra-rater reliability.[6] [32] [60] [61] [63] Vitally, the MCA codes AC, AI, and AB are sufficiently stable over time to be used in group studies of discourse in aphasia.[63] Recently, MCA checklists were developed using 92 control AphasiaBank speakers across the lifespan as a normative sample.[6] [64] Scoring information regarding specific lexical usage of these controls, alternative syntactic structures, and disallowed lexical items or structures are provided in appendices to facilitate reliable scoring when used in conjunction with Appendix A of Nicholas and Brookshire.[32] Normative control data, normative PWA data, comparisons between controls and all PWAs, and comparisons of aphasia subtypes to controls have been reported using these checklists.[6] [57] [64] Although MCA may not have been originally developed with non-transcription based analysis in mind, these tools may allow clinicians to reliably use it in this manner.

Core Lexicon

Core lexicon (CoreLex) analysis is used to investigate the typicality of words in discourse[4] [65] [66] [67] which researchers have shown to be correlated with both word-level and utterance-level discourse performance.[4] [65] For example, during Sandwich, a healthy control might use lexical items such as “bread” and “spread,” whereas an individual with aphasia might use lexical items such as “bun” and “push.” While the gist of the message might still be communicated, less typical lexical items may impact conversational partners' perceptions of the individual and may be more cognitively taxing for both partners. This issue includes a tutorial on CoreLex[51] which reviews how core lexicons are developed and might be applied by clinicians, and a compendium of CoreLex checklists for various stimuli.[68]

CoreLex is scored by assigning one point for each checklist item produced by a participant. Synonyms of CoreLex items are not given points, but inflected CoreLex items are. For example, if a CoreLex checklist included the verb “run,” productions such as “jog,” “sprint,” and “race” would not receive credit, but productions of “ran” and “running” would. By nature, CoreLex does not require phonetic transcription, and is amenable to online scoring, as clinicians could simply check off each item that a client produces from a checklist.

Derived Efficiency Measures

Yorkston and Beukelman introduced a measure to capture how efficiently speakers convey information—content units per minute (CUs/min).[45] Later works expanded efficiency scores by relating other informativeness measures, such as CIUs, to time (i.e., CIUs/min) or output (i.e., %CIUs).[31] [69] We focus on efficiency over time, as such measures are “useful in examining the effort needed by the speaker to produce discourse and the consequent effort needed by the listener in receiving the information effectively.”[70] We gravitate toward measures that require less clinician work, and timing the length of discourse is certainly less effortful than counting total words, concepts, or other output to calculate a percentage. Furthermore, time efficiency measures have been just as informative as output efficiency measures. CIUs/min demonstrates sensitivity to group differences,[31] stability over time,[31] [63] and is recommended to document treatment-induced change, clinical decision making, and group research studies.[63] This measure has been used in treatment studies in stroke-induced aphasia,[62] [71] [72] [73] [74] PPA,[24] and group research in traumatic brain injury.[75] Impressively, CIUs/min corresponds to measures of social validity in aphasia outcomes research, such as listener perception and listener comfort.[62] [72]

Efficiency measures above the word level are less common—few studies have examined MCA efficiency, and none CoreLex efficiency. The earliest measures, and most closely related to main concepts, are CUs/min, where content units are defined as a grouping of information always expressed as a unit, and percent informative minimal discourse units (%IMDUs), where IMDUs are defined as an intelligible unit with a single unambiguous message, relevant and informative, with new information.[69] In addition, percentage of accurate/complete main concepts (%ACMC)[69] and AC/min (or ACMC/min) have been introduced.[60] AC/min has also been validated with subtests of standardized aphasia assessments in speakers of Cantonese.[60] In English language speakers, AC/min is sensitive to differences between healthy controls and persons with dementia or aphasia, but not to differences between persons with dementia and PWAs.[61] Derived efficiency measures that relate informativeness scores to the time needed to complete discourse tasks are reliable, sensitive, and socially valid. Given the contribution of efficiency measures to our understanding of discourse impairment and treatment effects, research extending time efficiency to potentially non-transcription based measures, such as MCA and CoreLex, is needed.

Purpose

Our purpose is to provide clinicians with additional evidence in three brief studies regarding the sensitivity of MCA, CoreLex, and efficiency to (1) identify differences between healthy controls and persons with stable aphasia, (2) identify differences between healthy controls and persons with PPA, and (3) measure treatment response in individuals with stable aphasia. This evidence will allow clinicians to judge the utility of these discourse measures for routine clinical practice. Combined with the resources provided by others in this issue,[49] [51] [68] [76] clinicians should be well positioned to begin using non-transcription based discourse analysis in their clinical practice.

Sensitivity to Differences between Controls and Persons with Stable Aphasia

Methods

Participants

We examined differences in CoreLex scores, CoreLex efficiency (CoreLex/min), and AC main concepts efficiency (AC/min) for the five semispontaneous AphasiaBank discourse tasks reported by Dalton and Richardson.[57] Transcripts with phonetic error coding from 133 healthy controls (60 males, 73 females) and 206 persons with stable aphasia (112 males, 94 females) were retrieved from the AphasiaBank database (see [Table 1]). Included in the PWA sample were 25 individuals not aphasic by WAB (NABW) with persistent communication difficulties despite scoring above the WAB cut-off. The PWAs also included 77 individuals with anomic, 46 individuals with Broca's, 41 individuals with conduction, and 17 individuals with Wernicke's aphasia.

Table 1 Demographic Data for Healthy Controls and Persons with Stable Aphasia from the AphasiaBank Database

		Control	PWA	NABW	Anomic	Broca's	Conduction	Wernicke's
Age	Mean	64.5	61.1	60.6	62.1	56.2	63.1	65.9
	(SD)	(17.9)	(12.3)	(13.9)	(11.3)	(11.8)	(12.8)	(11.9)
Sex	Mean	60 males	112 males	8 males	43 males	30 males	20 males	11 males
	(SD)	73 females	94 females	17 females	34 females	16 females	21 females	6 females
Race/ Ethnicity		129 white 2 Hispanic/ Latino	181 white 5 Hispanic/ Latino 16 African American 1 mixed 1 Native American 2 Asian	22 white 2 Hispanic/ Latino 1 African American	73 white 1 Hispanic/ Latino 3 African American	36 white 2 Hispanic/ Latino 6 African American 1 mixed 1 Asian	36 white 3 African American 1 Native American 1 Asian	14 white 3 African American
Education	Mean	15.3	15.4	16	15.7	14.7	15.5	15.3
	(SD)	(2.4)	(2.8)	(3)	(2.7)	(2.6)	(3.2)	(2.4)
WAB-R-AQ	Mean		74.1	96.5	85.1	54.1	71.1	52.1
	(SD)		(18.2)	(1.8)	(6.8)	(14.6)	(8.9)	(14)

Abbreviation: WAB-R-AQ, Western Aphasia Battery—Revised Aphasia Quotient.

Data Analysis

All transcripts had previously been scored by the authors for main concept production using the Richardson and Dalton checklists.[6] [64] They were scored for CoreLex production according to the checklists reported by Dalton and Richardson[65] and Dalton et al.[68] Discourse sample lengths for efficiency calculations were retrieved using the EVAL tool in the Computerized Language ANalysis (CLAN) software available through AphasiaBank.

Upon examination, data were generally nonnormally distributed with heterogeneous shapes. Therefore, nonparametric median tests were used to compare healthy controls to PWAs and healthy controls to individuals with each aphasia subtype. Holm-Bonferroni corrections for multiple comparisons were applied to each variable separately. [Supplementary Material] reports descriptive statistics and normative data underlying these comparisons.

Results

MCA Codes and Scores

Previously, main concept composite scores (using the modified formula)[6] and the number of AC productions were significantly lower in PWAs than in healthy controls for all tasks.[57] PWAs produced more statements judged AI and AB for all tasks, more IC statements during Broken Window and Cat Rescue, and more II statements during Cinderella and Sandwich than in healthy controls. Differences between controls and PWAs with specific subtypes varied, with fewer significant differences between controls and individuals NABW than other subtypes. Overall, results indicated that MCA was sensitive to group differences at even the subtype level, depending on the discourse task.[57]

AC Efficiency

AC/min is reduced in PWAs compared with healthy controls ($p < 0.001$) for all tasks. When examining differences between subtypes and healthy controls, individuals NABW produced fewer ACs/min than controls during Broken Window ($p = 0.002$), Cat Rescue ($p = 0.001$), and Cinderella ($p = 0.006$). In contrast, individuals with anomic, Broca's, conduction, and Wernicke's aphasia produced fewer ACs/min than controls during all tasks ($p < 0.001$).

CoreLex Scores

For all tasks, healthy controls produced more CoreLex items than PWAs ($p < 0.001$ for all). When examining CoreLex production by subtype, individuals with anomic, conduction, Wernicke's, and Broca's aphasia produced fewer CoreLex items than controls for each task ($p < 0.005$ for all). Individuals NABW produced fewer CoreLex items than healthy controls during Cinderella ($p < 0.001$).

CoreLex Efficiency

PWAs produced fewer CoreLex/min than healthy controls for all tasks ($p < 0.001$ for all). Individuals with anomic, conduction, and Broca's aphasia produced fewer CoreLex/min than controls during Broken Window, Refused Umbrella, and Cat Rescue ($p < 0.001$). Individuals with Wernicke's aphasia produced fewer CoreLex/min during Broken Window, Cat Rescue, and Sandwich ($p < 0.005$). No significant differences in CoreLex/min were observed for individuals NABW compared with healthy controls.

Sensitivity to Differences between Controls and Persons with Progressive Aphasia

Methods

Participants

We examined the discourse of individuals with PPA to determine if AC main concepts, AC/min, CoreLex, and CoreLex/min were sensitive to changes in this population. PPA is a neurodegenerative condition where speech and language deficits are the initial and most prominent symptoms. There are three recognized variants of PPA, which have distinctive speech-language features, areas of atrophy, and underlying neuropathology. Because PPA and its variants may be unfamiliar to many clinicians and researchers, we provide a brief review here.

The nonfluent/agrammatic variant of PPA (nfvPPA) is characterized by agrammatic and/or motor speech impairment (commonly apraxia but may include dysarthria) with relative sparing of semantic memory.[77] [78] Associated areas of atrophy for nfvPPA include left frontoinsular regions and the supplementary motor area.[77] [78] The semantic variant of PPA (svPPA) is characterized by marked naming impairment and loss of semantic memory and object knowledge. Individuals with svPPA typically have fluent, empty speech and intact syntactic structures.[77] SvPPA is associated with bilateral anterior temporal lobe atrophy, generally with left-sided atrophy greater than right.[77] [79] The logopenic variant of PPA (lvPPA) is characterized by repetition and naming impairment and phonological paraphasia (speech sound errors without distortion). In the early stage, repetition and naming impairments may be mild, such that impairments may only be seen on complex or generative tasks (e.g., spontaneous speech, longer or nonmeaningful sentences). Individuals with lvPPA may have word-finding pauses, but speech is typically described as fluent based on the absence of motor speech impairment and agrammatism.[16] [77] Syntactic structures and utterance length are relatively spared.[3] [77] LvPPA is typically associated with atrophy in the left temporoparietal regions.[80]

Seventeen individuals with PPA (6 nfvPPA, 1 svPPA, 10 lvPPA) were included. Differential diagnosis of PPA variant was determined following the current consensus criteria guidelines.[77] These individuals had completed the AphasiaBank semispontaneous speech tasks as part of an extensive speech, language, and cognitive battery designed to allow differential diagnosis of PPA versus other neurocognitive disorders and identify the PPA variant most consistent with their behavioral profile. Healthy control transcripts of 91 individuals, aged 60 years and older, from the previously scored AphasiaBank sample were used for comparison ([Table 2]).

Table 2

Demographic Data for Healthy AphasiaBank Controls and Persons with Primary Progressive Aphasia

		Control	PPA	lvPPA	nfvPPA
Age	Mean (SD)	74.9 (6.9)	73.2 (9)	72.9 (8.1)	76.7 (8.3)
Sex		44 males 47 females	10 males 7 females	6 males 4 females	3 males 3 females
Race/Ethnicity		90 white 1 Hispanic/Latino	16 white 1 mixed	9 white 1 mixed	6 white

Education	Mean (SD)	15.2 (2.3)	17[a] (2.5)	18.2 (2.2)	14.8[a] (1.8)
MMSE (max 30)	Mean (SD)	–	21.3[b] (4.1)	21.7 (3)	19.3[b] (6.1)
BNT-SF (max 15)	Mean (SD)	–	10.6[b] (4.2)	10.4 (3.9)	13.3[b] (2.1)

Abbreviations: BNT-SF, Boston Naming Test Short Form; lvPPA, logopenic variant of PPA; MMSE, Mini-mental State Exam; nfvPPA, nonfluent/agrammatic variant of PPA; PPA, primary progressive aphasia.

[a] *One participant did not report education, so is not included in these averages.*

[b] *Two participants with nonfluent PPA did not complete the MMSE and BNT-SF, so are not included in these averages.*

Data Analysis

Discourse samples were orthographically transcribed in a word document but did not undergo a formal transcription procedure. AC main concepts and CoreLex checklists were scored as described earlier. Only AC main concepts were examined here due to the small sample size, as well as previous research indicating AC main concepts are sensitive to changes in other neurocognitive disorders.[61] Lengths of samples were retrieved from video recordings of the assessment by trained graduate research assistants in the third author's laboratory.

Upon examination, data were generally nonnormally distributed with heterogeneous shapes. Therefore, nonparametric median tests were used to compare performance between healthy controls and persons with PPA, as well as between all healthy controls and individuals with either logopenic or nonfluent variant PPA (svPPA was not examined separately because there was only one individual with svPPA in the data). Corrections for multiple comparisons were not used given the exploratory nature of the analysis. [Supplementary Material] reports descriptive statistics and normative data underlying these comparisons.

Results

AC Main Concepts

Individuals with PPA produced fewer AC main concepts than controls for all tasks ($p < 0.01$ for all). When examining each PPA variant separately, individuals with lvPPA produced fewer AC main concepts than controls during Cat Rescue ($p = 0.018$), Cinderella ($p = 0.007$), and Sandwich ($p = 0.007$). Individuals with nfvPPA produced fewer AC main concepts than healthy controls during Sandwich ($p = 0.048$).

AC Efficiency

Individuals with PPA produced fewer AC/min than controls for all tasks ($p < 0.01$ for all). Individuals with lvPPA were less efficient than healthy controls during Cat Rescue ($p = 0.021$), Cinderella ($p = 0.021$), Refused Umbrella ($p = 0.021$), and Sandwich ($p = 0.025$). Individuals with nfvPPA were less efficient during Cat Rescue ($p = 0.037$) and Cinderella ($p = 0.037$).

CoreLex Scores

Individuals with PPA produced fewer CoreLex items than controls for all tasks ($p < 0.05$ for all). During Cinderella, individuals with lvPPA ($p = 0.006$) and individuals with nfvPPA ($p = 0.047$) produced fewer CoreLex items than controls.

CoreLex Efficiency

Individuals with PPA produced fewer CoreLex/min than controls for all tasks ($p < 0.05$ for all). Individuals with lvPPA produced fewer CoreLex/min than controls during Cat Rescue ($p = 0.022$), Cinderella ($p = 0.003$), and Refused Umbrella ($p = 0.022$). Individuals with nfvPPA produced fewer CoreLex/min than controls during Cinderella ($p = 0.038$).

Preliminary Treatment Response in Persons with Chronic Stroke-Induced Aphasia

Methods

Participants

These data are from a study investigating the impact of brain stimulation (transcranial direct current stimulation) paired with behavioral therapy on speech-language outcomes. Participants completed 15 hours (1 hour/day) of semantic feature analysis and phonological component analysis (SFA/PCA) therapy across 15 consecutive weekdays (time split evenly between the two analysis approaches). Therapy was administered simultaneously with 30 minutes of active or sham brain stimulation, depending on randomization. The primary outcome was change in naming of trained items, but Broken Window, Refused Umbrella, and Cat Rescue discourse samples were elicited as a secondary outcome.

Enrollment is ongoing, but here we present preliminary data from 15 participants (4 NABW, 3 anomic, 3 conduction, 4 Broca's, and 1 transcortical sensory). We include participants who had been randomized into both brain stimulation groups, as all participants were expected to benefit from the behavioral therapy, and no detrimental effects were expected from the brain stimulation conditions. See [Table 3] for complete demographic information.

Table 3

Demographic Data for Individuals with Chronic Stroke-Induced Aphasia Participating in a Treatment Study Investigating the Impact of Brain Stimulation Paired with Speech-Language Therapy

	PWA
Age	55.8 (15.6)
Sex	12 males 3 females
Race/Ethnicity	10 white 5 Hispanic/Latino
Education	14 (3.2)
WAB-R-AQ	76.8 (19.2)

Abbreviation: WAB-R-AQ, Western Aphasia Battery—Revised Aphasia Quotient.

Data Analysis

Participant's discourse samples were scored using an orthographic transcript and/or video recording. Scores were collapsed across the three tasks, as these short samples are less stable than longer samples but combining tasks improve stability.[9] Difference scores for each measure were calculated as posttreatment minus pretreatment. A one-tailed, one sample *t*-test was used to determine if difference scores were greater than zero, since the hypothesis for this study was that treatment would improve language outcomes. Given the low power to detect effects in this preliminary analysis, effect sizes (Cohen's *d*) were calculated to examine the magnitude of therapy-induced changes in discourse. [Supplementary Material] includes descriptive statistics for these comparisons.

Results

This investigation revealed a significant treatment effect for an untrained discourse task, demonstrating the generalization potential of combined SFA/PCA therapy. There was increased production of AC main concepts ($t = 2.125$, $p = 0.026$) with a medium effect size ($d = 0.54$). Effect sizes for other measures showed a small treatment effect for change in CoreLex ($d = 0.35$) and AC/min ($d = 0.31$).

General Discussion

Stable Aphasia

Here we establish the first norms for AC/min, CoreLex, and CoreLex/min for healthy controls and PWAs for AphasiaBank semispontaneous speech tasks, which can be used in conjunction with previously established norms and checklists to obtain clinically informative discourse performance. Previous reports have shown the sensitivity of MCA to differences between healthy controls and PWAs.[32] [56] [57] [60] [61] Importantly, MC composite scores, MC attempts, and the number of AB and AC codes correlate with WAB aphasia quotient (AQ), suggesting that these index the overall severity of language impairment, while the number of error codes (AI, IC, and II) did not correlate with WAB-AQ, suggesting they index some aspect of language impairment other than overall severity.[57] Therefore, MC scores and codes can be used for both characterizing the severity of language impairments as well as specific deficits that may be contributing to communication difficulties.

AC efficiency, CoreLex scores, and CoreLex efficiency can also be used to differentiate between healthy controls and PWAs. Crucially, AC efficiency and CoreLex scores are sensitive enough to detect differences between healthy controls and those who score above the cut-off for a diagnosis of aphasia, consistent with Fromm and colleagues.[58] Combined with previous research reporting changes in similar informativeness measures in response to treatment,[81] [82] [83] [84] [85] and their relationship to measures of quality of life,[62] [81] these results suggest that MCA, CoreLex, and derived efficiency scores are informative discourse measures with standardized administration and accompanying normative data[6] [32] [40] [56] [57] [59] [60] [61] [64] [65] that may be appropriate for use as primary clinical outcomes as defined by Brady et al.[2]

Primary Progressive Aphasia

This is the first study to examine discourse informativeness, typicality, and efficiency in adults with PPA. Individuals with PPA produce discourse that is less informative and efficient, with fewer typical lexical items than healthy controls. Our results demonstrate that changes in discourse are apparent by the time patients and their families begin to seek a medical diagnosis or are referred to a SLP for evaluation and treatment. Due to the progressive nature of this disease, it can be expected that deficits will continue to become more evident, and these measures may be even more sensitive at later stages. As research seeks to improve our understanding of how language difficulties progress and how speech-language treatment interrupts that progression, sensitive measures of discourse performance that relate meaningfully to everyday communicative functions will be critical.

Although we did not compare the performance of individuals with lvPPA and nfvPPA to each other, we did see different patterns when we compared each group to healthy controls. Individuals with lvPPA had a larger number of significant differences when compared with controls than individuals with nfvPPA. While we were not able to examine individuals with svPPA, it might be expected that they would show worse performance on MCA, and especially CoreLex, due to the nature of their deficits. Individuals with nfvPPA may have the greatest protection against loss of informativeness, at least until the later stages of the disease, particularly if they present with a greater motor speech impairment and less agrammatism. The analysis conducted here cannot speak directly to these relationships, but further research is warranted. Finally, while we examined production of AC main concepts due to the limited sample size and previous research,[61] future investigations should examine all main concept codes and scores in a larger sample of PPA with improved distribution of PPA variants.

Preliminary Treatment Response in Chronic Aphasia

This preliminary investigation shows that MCA, CoreLex, and efficiency may be sensitive to changes as a result of speech-language therapy. Encouragingly, these results are seen in response to a behavioral therapy that did not focus directly on discourse, rather on improving naming through analysis of semantic and phonological features. This suggests both that improvements gained during combined SFA/PCA therapy may generalize to more functional communication, and that MCA, CoreLex, and efficiency may be sensitive enough to identify changes with relatively low power. Given these promising results, future research should continue to investigate the utility of these measures for determining treatment response.

Future Directions

Before full clinical utilization of these measures can be achieved, further research is needed. In particular, an investigation that directly compares the accuracy and reliability of MCA and CoreLex from formal transcripts, orthographic transcripts, and video or audio is needed. In the first study utilizing AphasiaBank data, orthographic transcripts with phonetic error coding were used. However, in the latter two studies, we scored discourse of individuals with PPA and pretest/posttest discourse of chronic, stroke-induced PWAs using a combination of audio/video recordings and orthographic transcripts, indicating that non-transcription based scoring of these measures is feasible, if good accuracy and reliability can be confirmed. In addition, investigations confirming the stability of these measures over time are needed. While some preliminary research indicates that various discourse measures and tasks have adequate stability,[9] [63] these specific measures and tasks should undergo that investigation as well.

Conclusion

Shifting foci in clinical decision making and research necessitates the use of outcome measures that relate to meaningful aspects of everyday, functional communication. Calls for improvements in discourse measures (e.g., publication of normative data, reporting of psychometric properties, etc.) are becoming more frequent and urgent.[11] [14] We have demonstrated the utility of MCA, CoreLex, and efficiency measures to differentiate individuals with stable and progressive aphasia from healthy controls, as well as preliminary results of these measures for treatment outcomes. We have provided additional tools that can be used alongside MCA checklists and norms,[6] [57] [64] CoreLex checklists and norms,[65] [68] and the AphasiaBank protocol[47] [48] to sensitively and reliably assess discourse. These measures hold the potential for non-transcription based implementation, addressing one of the more intractable barriers to clinical implementation of discourse outcomes. It is our hope that researchers and practicing clinicians will consider implementation of these or similar measures in the future to track functional communication changes in their clients.

Conflict of Interest

None declared.

Supplementary Material

Supplementary Material

References

1 Worrall L, Sherratt S, Rogers P. , et al. What people with aphasia want: their goals according to the ICF. Aphasiology 2011; 25 (03) 309-322

2 Brady MC, Kelly H, Godwin J, Enderby P, Campbell P. Speech and language therapy for aphasia following stroke. Cochrane database of Syst Rev 2016; (06) 1-314

3 Ash S, Grossman M. Why study connected speech production. In: Willems R. , ed. Cognitive Neuroscience of Natural Language Use. Cambridge, UK: Cambridge University Press; 2015: 29-58

4 Kim H, Kintz S, Zelnosky K, Wright HH. Measuring word retrieval in narrative discourse: core lexicon in aphasia. Int J Lang Commun Disord 2019; 54 (01) 62-78

5 Kong AP, Linnik A, Law SP, Shum WW. Measuring discourse coherence in anomic aphasia using Rhetorical Structure Theory. Int J Speech Lang Pathol 2018; 20 (04) 406-421

6 Richardson JD, Dalton SG. Main concepts for three different discourse tasks in a large non-clinical sample. Aphasiology 2016; 30 (01) 45-73

7 Whitworth A, Claessen M, Leitão S, Webster J. Beyond narrative: Is there an implicit structure to the way in which adults organise their discourse?. Clin Linguist Phon 2015; 29 (06) 455-481

8 Wright HH, Capilouto GJ, Koutsoftas A. Evaluating measures of global coherence ability in stories in adults. Int J Lang Commun Disord 2013; 48 (03) 249-256

9 Boyle M. Stability of word-retrieval errors with the AphasiaBank stimuli. Am J Speech Lang Pathol 2015; 24 (04) S953-S960

10 Leaman MC, Edmonds LA. Revisiting the correct information unit: measuring informativeness in unstructured conversations in people with aphasia. Am J Speech Lang Pathol 2019; 28 (03) 1099-1114
11 Pritchard M, Hilari K, Cocks N, Dipper L. Psychometric properties of discourse measures in aphasia: acceptability, reliability, and validity. Int J Lang Commun Disord 2018; 53 (06) 1078-1093
12 Armstrong E, Bryant L, Ferguson A, Simmons-Mackie N. Approaches to assessment and treatment of everyday talk in aphasia. In: Papathanasiou I, Coppens P, Potagas C. , eds. Aphasia and Related Neurogenic Communication Disorders. Burlington, MA: Jones & Bartlett Learning; 2016: 269-285
13 Bryant L, Ferguson A, Spencer E. Linguistic analysis of discourse in aphasia: a review of the literature. Clin Linguist Phon 2016; 30 (07) 489-518
14 Linnik A, Bastiaanse R, Höhle B. Discourse production in aphasia: a current review of theoretical and methodological challenges. Aphasiology 2016; 30 (07) 765-800
15 Ash S, Moore P, Antani S, McCawley G, Work M, Grossman M. Trying to tell a tale: discourse impairments in progressive aphasia and frontotemporal dementia. Neurology 2006; 66 (09) 1405-1413
16 Ash S, Evans E, O'Shea J. , et al. Differentiating primary progressive aphasias in a brief sample of connected speech. Neurology 2013; 81 (04) 329-336
17 Dalton SGH, Shultz C, Henry ML, Hillis AE, Richardson JD. Describing phonological paraphasias in three variants of primary progressive aphasia. Am J Speech Lang Pathol 2018; 27 (1S): 336-349
18 Fraser KC, Meltzer JA, Graham NL. , et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex 2014; 55: 43-60
19 Sajjadi SA, Patterson K, Tomek M, Nestor PJ. Abnormalities of connected speech in the non-semantic variants of primary progressive aphasia. Aphasiology 2012; 26 (10) 1219-1237
20 Hird K, Brown R, Kirsner K. Stability of lexical deficits in primary progressive aphasia: evidence from natural language. Brain Lang 2006; 99 (1–2): 147-148
21 Knibb JA, Woollams AM, Hodges JR, Patterson K. Making sense of progressive non-fluent aphasia: an analysis of conversational speech. Brain 2009; 132 (Pt 10): 2734-2746
22 Taylor C, Croot K, Power E, Savage SA, Hodges JR, Togher L. Trouble and repair during conversations of people with primary progressive aphasia. Aphasiology 2014; 28 (8–9): 1069-1091
23 Croot K, Taylor C, Abel S. , et al. Measuring gains in connected speech following treatment for word retrieval: a study with two participants with primary progressive aphasia. Aphasiology 2014; 29 (11) 1265-1288
24 Whitworth A, Cartwright J, Beales A, Leitão S, Panegyres PK, Kane R. Taking words to a new level: a preliminary investigation of discourse intervention in primary progressive aphasia. Aphasiology 2017; 32 (11) 1284-1309
25 Kagan A, Simmons-Mackie N. Beginning with the end: outcome-driven assessment and intervention with life participation in mind. Top Lang Disord 2007; 27 (04) 309-317
26 Simmons-Mackie N, Threats TT, Kagan A. Outcome assessment in aphasia: a survey. J Commun Disord 2005; 38 (01) 1-27
27 Goodglass H, Kaplan E, Brand S, Barresi B. The Boston Diagnostic Aphasia Examination (BDAE). Philadelphia, PA: Lippincott Williams; 2000
28 Kertesz A. Western Aphasia Battery - Revised. San Antonio: Pearson; 2007

29 Prins R, Bastiaanse R. Analyzing the spontaneous speech of aphasic speakers. Aphasiology 2004; 18 (12) 1075-1091
30 Wright HH, Fergadiotis G. Conceptualising and measuring working memory and its relationship to aphasia. Aphasiology 2012; 26 (3-4): 258-278
31 Nicholas LE, Brookshire RH. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. J Speech Hear Res 1993; 36 (02) 338-350
32 Nicholas LE, Brookshire RH. Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. J Speech Hear Res 1995; 38 (01) 145-156
33 Mayer J, Murray L. Functional measures of naming in aphasia: word retrieval in confrontation naming versus connected speech. Aphasiology 2003; 17 (05) 481-497
34 Dodd B. Evidence-based practice and speech-language pathology: strengths, weaknesses, opportunities and threats. Folia Phoniatr Logop 2007; 59 (03) 118-129
35 Jette DU, Bacon K, Batty C. , et al. Evidence-based practice: beliefs, attitudes, knowledge, and behaviors of physical therapists. Phys Ther 2003; 83 (09) 786-805
36 Dysart AM, Tomlin GS. Factors related to evidence-based practice among U.S. occupational therapy clinicians. Am J Occup Ther 2002; 56 (03) 275-284
37 Nail-Chiwetalu BJ, Ratner NB. Information literacy for speech-language pathologists: a key to evidence-based practice. Lang Speech Hear Serv Sch 2006; 37 (03) 157-167
38 Brookshire RH, Nicholas LE. Speech sample size and test-retest stability of connected speech measures for adults with aphasia. J Speech Hear Res 1994; 37 (02) 399-407
39 Capilouto G, Wright HH, Wagovich SA. CIU and main event analyses of the structured discourse of older and younger adults. J Commun Disord 2005; 38 (06) 431-444
40 Nicholas LE, Brookshire RH. A system for scoring main concepts in the discourse of non-brain-damaged and aphasic speakers. Clinical Aphasiology Archive 1993; 21: 87-99
41 Wright HH, Shisler RJ. Working memory in aphasia: theory, measures, and clinical implications. Am J Speech Lang Pathol 2005; 14 (02) 107-118
42 Wright HH, Koutsoftas AD, Capilouto GJ, Fergadiotis G. Global coherence in younger and older adults: influence of cognitive processes and discourse type. Neuropsychol Dev Cogn B Aging Neuropsychol Cogn 2014; 21 (02) 174-196
43 Fergadiotis G, Wright HH, Capilouto GJ. Productive vocabulary across discourse types. Aphasiology 2011; 25 (10) 1261-1278
44 Pratt MW, Boyes C, Robins S, Manchester J. Telling tales: aging, working memory, and the narrative cohesion of story retellings. Dev Psychol 1989; 25 (04) 628-635
45 Yorkston KM, Beukelman DR. An analysis of connected speech samples of aphasic and normal speakers. J Speech Hear Disord 1980; 45 (01) 27-36
46 Mackenzie C, Brady M, Norrie J, Poedjianto N. Picture description in neurologically normal adults: Concepts and topic coherence. Aphasiology 2007; 21 (3–4): 340-354
47 Macwhinney B, Fromm D, Forbes M, Holland A. AphasiaBank: methods for studying discourse. Aphasiology 2011; 25 (11) 1286-1307
48 Forbes MM, Fromm D, Macwhinney B. AphasiaBank: a resource for clinicians. Semin Speech Lang 2012; 33 (03) 217-222

49 Fromm D, Forbes M, Holland A, MacWhinney B. Using AphasiaBank for discourse assessment. Semin Speech Lang 2020; 41: 10-19
50 Duncan EA, Murray J. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. BMC Health Serv Res 2012; 12 (01) 96
51 Kim H, Wright HH. A tutorial on core lexicon: development, use, and application. Semin Speech Lang 2020; 41: 20-31
52 Armstrong L, Brady M, Mackenzie C, Norrie J. Transcription-less analysis of aphasic discourse: a clinicians dream or a possibility?. Aphasiology 2007; 21 (3–4): 355-374
53 Boles L, Bombard T. Conversational discourse analysis: appropriate and useful sample sizes. Aphasiology 1998; 12 (7–8): 547-560
54 Olness GS, Gyger J, Thomas K. Analysis of narrative functionality: toward evidence-based approaches in managed care settings. Semin Speech Lang 2012; 33 (01) 55-67
55 Bryant L, Spencer E, Ferguson A. Clinical use of linguistic discourse analysis for the assessment of language in aphasia. Aphasiology 2016; 31 (10) 1105-1126
56 Kong AP. The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: a preliminary report. J Commun Disord 2009; 42 (06) 442-464
57 Dalton SGH, Richardson JD. A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. Am J Speech Lang Pathol 2019; 28 (1S): 293-320
58 Fromm D, Forbes M, Holland A, Dalton SG, Richardson J, MacWhinney B. Discourse characteristics in aphasia beyond the western aphasia battery cutoff. Am J Speech Lang Pathol 2017; 26 (03) 762-768
59 Hameister I, Nickels L. The cat in the tree – using picture descriptions to inform our understanding of conceptualisation in aphasia. Lang Cogn Neurosci 2018; 33 (10) 1296-1314
60 Kong AP. The main concept analysis in Cantonese aphasic oral discourse: external validation and monitoring chronic aphasia. J Speech Lang Hear Res 2011; 54 (01) 148-159
61 Kong AP, Whiteside J, Bargmann P. The main concept analysis: validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. Logoped Phoniatr Vocol 2016; 41 (03) 129-141
62 Ross K, Wertz RT. Comparison of impairment and disability measures for assessing severity of, and improvement in, aphasia. Aphasiology 1999; 13 (02) 113-124
63 Boyle M. Test-retest stability of word retrieval in aphasic discourse. J Speech Lang Hear Res 2014; 57 (03) 966-978
64 Richardson JD, Dalton SGH. Main concepts for two picture description tasks: an addition to Richardson and Dalton, 2016. Aphasiology 2019; 1-18 . [Epub ahead of print]
65 Dalton SG, Richardson JD. Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. Am J Speech Lang Pathol 2015; 24 (04) S923-S938
66 Fromm DA, Forbes M, Holland A, MacWhinney B. PWAs and PBJs: language for describing a simple procedure. [Abstract]. Arch Clinical Aphasiology 2013. Available at: http://aphasiology.pitt.edu/id/eprint/2491 . Accessed October 10, 2019
67 MacWhinney B, Fromm D, Holland A, Forbes M, Wright H. Automated analysis of the Cinderella story. Aphasiology 2010; 24 (6-8): 856-868

68 Dalton SGH, Kim H, Richardson JD, Wright HH. A compendium of core lexicon checklists. Semin Speech Lang 2020; 41: 45-60

69 Doyle PJ, Goda AJ, Spencer KA. The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. Am J Speech Lang Pathol 1995; 4 (04) 130-134

70 Armstrong E. Aphasic discourse analysis: the story so far. Aphasiology 2000; 14 (09) 875-892

71 Ballard KJ, Thompson CK. Treatment and generalization of complex sentence production in agrammatism. J Speech Lang Hear Res 1999; 42 (03) 690-707

72 Jacobs BJ. Social validity of changes in informativeness and efficiency of aphasic discourse following linguistic specific treatment (LST). Brain Lang 2001; 78 (01) 115-127

73 Mozeiko J, Coelho CA, Myers EB. The role of intensity in constraint-induced language therapy for people with chronic aphasia. Aphasiology 2016; 30 (04) 339-363

74 Savage MC, Donovan NJ. Comparing linguistic complexity and efficiency in conversations from stimulation and conversation therapy in aphasia. Int J Lang Commun Disord 2017; 52 (01) 21-29

75 Carlomagno S, Giannotti S, Vorano L, Marini A. Discourse information content in non-aphasic adults with brain injury: a pilot study. Brain Inj 2011; 25 (10) 1010-1018

76 Boyle M. Choosing discourse outcome measures to assess clinical change. Semin Speech Lang 2020; 41: 1-9

77 Gorno-Tempini ML, Hillis AE, Weintraub S. , et al. Classification of primary progressive aphasia and its variants. Neurology 2011; 76 (11) 1006-1014

78 Caso F, Mandelli ML, Henry M. , et al. In vivo signatures of nonfluent/agrammatic primary progressive aphasia caused by FTLD pathology. Neurology 2014; 82 (03) 239-247

79 Collins JA, Montal V, Hochberg D. , et al. Focal temporal pole atrophy and network degeneration in semantic variant primary progressive aphasia. Brain 2017; 140 (02) 457-471

80 Rohrer JD, Caso F, Mahoney C. , et al. Patterns of longitudinal brain atrophy in the logopenic variant of primary progressive aphasia. Brain Lang 2013; 127 (02) 121-126

81 Cupit J, Rochon E, Leonard C, Laird L. Social validation as a measure of improvement after aphasia treatment: Its usefulness and influencing factors. Aphasiology 2010; 24 (11) 1486-1500

82 Avent J, Austermann S. Reciprocal scaffolding: a context for communication treatment in aphasia. Aphasiology 2003; 17 (04) 397-404

83 Albright E, Purves B. Exploring SentenceShaper™: treatment and augmentative possibilities. Aphasiology 2008; 22 (7–8): 741-752

84 Coelho CA, McHugh RE, Boyle M. Semantic feature analysis as a treatment for aphasic dysnomia: a replication. Aphasiology 2000; 14 (02) 133-142

85 Stark JA. Content analysis of the fairy tale Cinderella – a longitudinal single-case study of narrative production: "From rags to riches. Aphasiology 2010; 24 (6–8): 709-724

Address for correspondence

Sarah Grace Hudspeth Dalton, Ph.D., CCC-SLP
Department of Speech Pathology and Audiology, Marquette University
604 N 16th St., Milwaukee, WI 53233

Email: sarahgrace.dalton@marquette.edu