A Tutorial on Core Lexicon: Development, Use, and Application

Hana Kim, M.A.[1] and Heather Harris Wright, Ph.D.[1]

ABSTRACT

Evidence suggests that discourse-level assessment in aphasia should be implemented within clinical settings. However, existing discourse measures that are time and labor intensive in process prevent speech-language pathologists from applying such measures to their clinical practices. This article provides an overview of a lexicon-based analysis (core lexicon measure) that recently was developed and investigated for clinical usability. A new approach to core lexicon measures provides a simple scoring method with short instructions, which may be practical and time efficient for assessment and management of persons with aphasia. The article concludes with suggestions for clinical application and implementation.

KEYWORDS: core lexicon, discourse analysis, aphasia

Learning Outcomes: As a result of this activity, the reader will be able to (1) explain issues that arise in the evaluation of discourse analysis in clinical settings; (2) discuss a novel, lexicon-based approach for measuring word retrieval ability at the discourse level; and (3) implement core lexicon measures for the assessment of PWA.

Over the past decade, perception of discourse outcome measures by clinicians and researchers has shifted, from that of a secondary measure to a primary measure in aphasia assessment.[1,2] Such change has ignited researchers' interest in the development of high-quality measures which are psychometrically robust.[3] Despite the conceptual advancement in discourse outcome measures, it is undeniable that clinical application and usability of theoretically well-established outcome measures have been overlooked. Maddy et al examined the extent to which clinicians have used discourse analysis in language assessment, finding a gap between clinicians' value of discourse analysis and their actual practice exists.[4] Specifically, speech-language pathologists (SLPs) understand the importance of discourse analysis for

[1]Department of Communication Sciences and Disorders, East Carolina University, Greenville, North Carolina.

Address for correspondence: Hana Kim, M.A., Department of Communication Sciences and Disorders, East Carolina University, 3310 Health Sciences Building, Greenville, NC 27834
(e-mail: only.hana.kim@gmail.com).

Advances in Discourse Assessment and Treatment; Guest Editor, Jessica D. Richardson, Ph.D., CCC-SLP.

Semin Speech Lang 2020;41:20–31.
DOI: https://doi.org/10.1055/s-0039-3400973.
ISSN 0734-0478.

evaluating patients' communicative exchanges. However, resource-intensive procedures for measuring discourse-level language performance hamper SLPs use of discourse measures in typical clinical settings. Bryant et al's study confirmed the barriers of discourse analysis in clinical settings.[5] In their survey, nearly half of the clinicians reported that they have *never* implemented discourse analysis during language assessment. They responded that the processes to elicit, transcribe, and analyze discourse samples are burdensome.

For many years, the key elements for successful, clinical outcome measures for discourse analysis have been discussed. The first point relevant to clinical use of discourse analysis concerns a cost–benefit analysis.[6–8] Clinicians generally provide assessment and treatment on the basis of a cost–benefit analysis for their patients.[6] It is doubtful that in this situation, existing discourse measures are effective and compelling means for assessing language performance in a clinical setting. The second issue concerns discourse elicitation techniques.[6,9–11] Researchers suggested selecting discourse elicitation techniques that best represent communicative exchanges or best fit with specific outcome measures while realizing the limited time in clinical settings; however, how well these tasks predict real-world communication is not well known. A third point concerns the time and training for completing discourse analysis.[7,12] A trained clinician generally requires more than four times the actual length of the discourse sample to complete only the transcription process.[13–15] This excludes the time required for training to reliably complete the analysis and completing the analysis, thus making many analyses impractical for use in clinical settings. The fourth issue concerns reliability of transcribing and segmenting discourse samples. Generally, discourse samples of persons with acquired neurogenic disorders include verbal output errors such as fillers and paraphasias, which reduce the accuracy of the transcription. The inaccurate transcriptions consequently lead SLPs to obtaining imprecise results and assessments. Following the completion of transcriptions, another preliminary stage in the discourse analysis is the identification of utterance structure relevant to specific measures. Discourse samples can be organized in a variety of ways. For example, c-unit[16] is one variation of T-unit,[17] and while roughly equivalent, it has some differences. Idea-unit developed by Kroll is another variation of t-unit, which was purported to be specific to the communicative nature of language samples.[18] National-unit[19] and information-unit[20] are also alternate methods of language segmentation. Microlinguistic level outcomes may be sensitive to the way we segment discourse samples,[7] as they are measured by dividing the total number of a specific utterance unit to avoid the impact of length of discourse samples. A fifth point relevant to clinical use of discourse analysis involves the lack of normative data to provide clinical guidance for interpreting results of the analyses.[8,9,21] Without normative data, clinicians are unable to interpret patients' performance, or further assess their linguistic changes following treatment.

Given the need for considering communication ability in persons with aphasia (PWA) and the limited application of current discourse methods in clinical settings, this review has three goals: (1) to introduce the recently developed concept—core lexicon measure as a clinician-friendly tool; (2) to inform SLPs how to use this measure for assessment; and (3) to demonstrate potential application of the core lexicon measure in clinical settings.

DEVELOPMENT OF CORE LEXICON MEASURES

In 2010, MacWhinney and colleagues introduced the use of TalkBank tools, a large shared database for aphasia research.[22] They indicated the importance of normative data as a comparison for SLPs and researchers to readily contrast the performance between cognitively healthy adults and clinical populations. They focused on illustrating a method for studying the patterns of lexical usage in structured narrative tasks. They demonstrated how the essential verb and noun lexicons required to deliver the Cinderella story were identified from language samples. The 10 most frequently occurring nouns and verbs produced by the two groups were compared. The researchers reported that six of the ten verbs were in common between the two groups, and the nouns produced by the PWA were not tightly linked to the Cinderella story as much as the

nouns produced by the controls. The PWA delivered the story using more general words, such as man, girl, home, and shoe, whereas the controls used precise target words such as prince, fairy, and godmother. Although the researchers did not provide further explanation regarding the underlying reason for these findings, it could be an indication of PWA's difficulty to retrieve target words in discourse. The same method was applied to a procedural task of how to make a peanut butter and jelly sandwich in a different study.[23] No difference in the top 10 essential noun items and two different verb items in the 10 essential verbs were found between the control and the PWA group.

Hudspeth et al further explored the Cinderella story by expanding the lexical options for generating a lexicon list, and first referenced the measure as core lexicon.[24] Unlike previous studies, for core lexicon lists, they selected lexical items produced by greater than 50% of the sampling cohort. Although they initially intended to include adjectives in their core lexicon list, the production of adjectives in language samples did not reach the criterion for lexical selection. As such, core verbs, core nouns, and an aggregated core lexicon list were generated, and then used to investigate if the core lexicon lists differentiate aphasia subtypes. Core verbs differed for the following groupings: persons with anomic aphasia produced more core verbs than persons with conduction and Broca's aphasia. Persons with Wernicke's aphasia produced more core verbs than persons with Broca's aphasia. Persons with anomic aphasia produced more core nouns than persons with Broca's and Wernicke's aphasia. Persons with conduction aphasia produced more core nouns than persons with Broca's and Wernicke's aphasia. When considering the combined core lexicon list, persons with Broca's aphasia produced less core lexicon items than persons with anomic aphasia and conduction aphasia. These findings suggest that separate core lexicon lists organized by word class may be more useful than a combined core lexicon list in that they demonstrate varying degrees of discrimination among aphasia subtypes.

Following the same criterion, Dalton and Richardson extracted all words produced for a sequential picture description task (Broken Window). Significant differences for number of core lexicon items were found between the PWA ($N = 92$) and control participants ($N = 166$). Further statistical testing demonstrated differences between aphasia subtypes and controls. The researchers also used main concept (MC) analysis—a measure of how accurately speakers deliver the gist of the narration—and found statistically significant correlations between core lexicon performance and MC scores. They concluded that performance based on the core lexicon measure might reflect concept-level discourse abilities, and that it may be related to PWA's ability to construct the content of the story.

In a recent study, age and word class was considered when developing core lexicon lists.[25] The researchers included 470 language samples collected from cognitively healthy adults for two narrative discourse tasks (*Good Dog Carl* [*GDC*][26] and *Picnic*[27]) and identified the 25 lexical items for each core lexicon list (nouns, verbs, adjectives, adverbs) among seven age groups (20s, 30s, 40s, 50s, 60s, 70s, and 80s). Eleven PWA were included to compare their performance and percent agreement for each core lexicon list was determined. Percent agreement was calculated by comparing the total number of items (25 items) within each list to the number of items produced by each PWA. Then, Spearman's correlation coefficient was computed between the percent agreement and the overall severity of aphasia as determined by the aphasia quotient (AQ) from the Western Aphasia Battery—Revised (WAB-R).[28] Significant correlations were found between percent of core verbs produced by the PWA and AQs. Percent of core verbs produced also differed between PWA based on fluency. The participants with fluent aphasia produced a greater percent of the core verbs than participants with nonfluent aphasia. In a subsequent study, the same group of researchers developed a 25-core function word list by using the same tasks and method.[29] Significant correlations were found between core function word agreement and aphasia severity as measured by the WAB-R. The researchers explained that the absence of function words in PWA's utterances might be related to their attempt to adapt to

impaired language processing. PWA tend to use an elliptical strategy with function words to compensate for their reduced cognitive sources,[30] resulting in reduced discourse production.

CLINICAL USE AND IMPLICATIONS OF CORE LEXICON MEASURES

Though only a handful of core lexicon studies have been conducted with PWA, advantages of using the core lexicon in clinical settings are apparent. First and foremost, core lexicon measures are devised to measure lexical usage based on normal language processing, aiming to provide a norm reference for clinical populations. Given that observed deficits seen in PWA reflect disruptions involved in normal processing, understanding where a PWA's performance situates in the continuum of normal performance is helpful for planning treatment. A second advantage is that core lexicon measures provide a checklist of the target lexicon items for a specific discourse elicitation task, not requiring the typical time-consuming activities of other discourse analyses (e.g., transcribing, segmenting discourse samples, training). Bypassing this discourse analysis preparatory work helps achieve error-free data and high-reliability across evaluators. Lastly, core lexicon measures do not provide lengthy guidelines that need to be mastered prior to scoring. Once clinicians are familiar with the core lexical item checklists, scoring can be completed with media (i.e., audio or video) files and without transcripts and potentially in real time.[31]

Although not all previous core lexicon studies considered different word classes, development of separate core lexicon lists by word class has ecological importance. For example, selective impairments of content words (nouns and verbs) and function words have been reported across aphasia subtypes, and the findings on dissociation patterns are mixed.[32–39] There are suggestions that there may be multiple factors that affect word retrieval performance, such as linguistic complexity and processing load. These factors are intimately and differentially related depending on word classes, and retrieval of all word classes is not equally impaired. In addition, it is surprising that relatively little attention has been devoted to modifiers as all word classes are considered to be unique elements bearing a semantic and/or syntactic role in discourse.[40–43] Regardless of how or why these selective impairments manifest in different aphasia subtypes, what matters most to SLPs is how to examine individually distinct profiles of language impairments in PWA. In clinical practice, standardized test batteries, such as the WAB-R[28] and the Boston Naming Test,[44] are the most common assessment tools used for evaluating severity and type of aphasia.[45–48] We are aware that such test batteries are constructed with little consideration of selective impairments of word classes in PWA's discourse processing. Therefore, multiple core lexicon measures by word class can provide a more complete understanding of PWA's specific lexical impairments in discourse, which will lead to clearer rationale for diagnosis and intervention plans.

Validity and Reliability of Core Lexicon Measures

A substantive issue in adapting a new language test to everyday clinical settings is whether the measure is valid and reliable for quantifying the intended linguistic behavior. The concern about using a new measure for clinical purposes can be addressed by statistically ensuring the accuracy of the measure and foundation of theoretical construct, which involves reliability and validity tests. There are three types of reliability associated with the quality of the test: internal consistency, test–retest stability, and inter-rater reliability.[49] Internal consistency represents the constancy of results across items, which is generally estimated using Cronbach's α.[50] Test–retest reliability reflects the stability of results across time. Inter-rater reliability is a measure of consistency between different examiners administering a test, which can be estimated by intraclass correlation coefficients (ICC).[51,52] Validity is equally important in that a new language measure that is reliable is not always found to be valid.[53] Validity has been investigated from four different perspectives relevant to aphasia language batteries: face validity, content validity, concurrent validity, and construct validity.[49] Face validity refers to the degree to which tests appear to measure what they were designed to measure. It is a

subjective judgement by test administrators. Content validity provides evidence about how well a test measures the domain of functions intended to be measured. Concurrent validity represents the relationship between the score on a test and scores on existing tests that are theoretically considered to index the same underlying behaviors. Lastly, construct validity pertains to the extent to which tests actually measure what they were intended to measure. Although studies and development of core lexicon measures are in their nascent stage, their reliability and validity have been comparatively well investigated.

Dalton and Richardson focused on formulating a broader picture of constructs of core lexicon measures.[54] To establish the construct validity of the core lexicon measure, the researchers used MC analysis, which is a measure of how accurately speakers deliver the gist of the narration. They hypothesized that PWA's difficulties to retrieve words required to deliver a narrative would impinge on their ability to deliver the gist of the narration. Statistically significant, positive correlations were found between the core lexicon measure and MC after collapsing across groups (controls and PWA; $r = 0.868, p < 0.001$) and after separating groups for PWA ($r = 0.738$, $p < 0.001$) and for the controls ($r = 0.630$, $p < 0.001$). Correlations were also significant and positive for all subgroups: anomic ($r = 0.710$, $p < 0.001$), Broca's ($r = 0.742$, $p < 0.001$), conduction ($r = 0.463, p < 0.001$), Wernicke's aphasia ($r = 0.707$, $p < 0.001$), and PWA who were not diagnosed as aphasia by WAB-R ($r = 0.519$, $p < 0.001$). The researchers concluded that the core lexicon measure may reflect decrements in information related to the content and structure of the narrative.

In a recent study of core lexicon measures with 11 PWA, Kim and Wright have yielded encouraging data related to concurrent validity and inter-rater reliability of their core lexicon measures by story task (*GDC*, *Picnic*).[55] Though the core lexicon measure was designed to provide information about the typicality of language use, it conceptually can be considered to index microlinguistic levels of language ability. Based on the examination by Dalton and Richardson,[54] the researchers hypothesized that performance on core lexicon and macrolinguistic measures would significantly correlate. As indices of microlinguistic measures, the researchers chose three different outcome measures: information units,[56,57] syntactic complexity,[58,59] and lexical diversity.[60–62] As indices of macrolinguistic measures, coherence and thematic units[63–65] were included. Results showed significant correlations among core lexicon measures and micro- and macrolinguistic indices, though different findings emerged depending on story tasks and word classes. More specifically, for *GDC*, significant correlations were found between core nouns and coherence ($r = 0.671$, $p < 0.05$), and thematic units ($r = 0.736$, $p < 0.05$); core adverbs and information units ($r = -763, p < 0.05$), and lexical diversity ($r = -0.661$, $p < 0.05$); and core function words and syntactic complexity ($r = 0.722$, $p < 0.05$). For *Picnic*, significant and positive correlations were found between core verbs and syntactic complexity ($r = 0.616, p < 0.05$), and lexical diversity ($r = 0.630, p < 0.05$); core nouns and coherence ($r = 0.654, p < 0.05$), thematic units ($r = 0.627$, $p < 0.05$), syntactic complexity ($r = 0.657$, $p < 0.05$), and lexical diversity ($r = 0.627, p < 0.05$); core adjectives and information units ($r = 0.636$, $p < 0.05$), and lexical diversity ($r = 0.701$, $p < 0.05$); and core function words and coherence ($r = 0.778, p < 0.01$), thematic units ($r = 0.634$, $p < 0.05$), syntactic complexity ($r = 0.803$, $p < 0.01$), and lexical diversity ($r = 0.824, p < 0.01$). We suggested that some core lexicon lists may reflect linguistic processes across different levels of discourse production. These findings served as the first indication that core lexicon measures may be a means of predicting lexical-semantic features in discourse.

Along with establishing concurrent validity, inter-rater reliability was investigated as another critical psychometric property to be approved prior to clinical use. To determine reliability, absolute-agreement ICC was calculated on scores among four raters, who have varying experience of discourse analysis, and determined the core lexicon scores for each sample. The raters were instructed to check the words from the core lexicon list when they heard them in the participant's stories. In an attempt to consider typical time available for clinicians in clinical settings to complete assessments, raters were able to listen to each story no more than two times for each list. Results indicated that all ICCs

were greater than 0.705, which is a cut-off of strong reliability suggested by Shrout and Fleiss's guidelines.[52] The researchers suggested that the core lexicon measure would be a viable option to reconcile ecological validity with clinical usability, given that only a one-time brief training session was provided.

Taken together, the aforementioned research findings provide empirical support for use of the core lexicon measure in clinical settings by demonstrating what core lexicon measures are purported to measure. The use of core lexicon measures permits clinical examinations of word retrieval ability not only at microlinguistic levels but also at microlinguistic levels. Moreover, as hypothesized by researchers that core lexicon measures would demonstrate higher reliability among multiple raters, core lexicon measures hold promise as a reliable measure. While these findings provide sufficient evidence to apply this measure in clinical practices, additional aspects of validity and reliability should be established to provide stronger evidence of the interpretation of test scores.

Guidelines for Scoring Lexical Usage in Discourse Using Core Lexicon Lists

Previous studies generating core lexicon lists demonstrate different principles that possibly affect clinical practices. Kim and colleagues' core lexicon lists were constructed to consider age-related differences in selecting lexical items, and differential performance by word class. Dalton and Richardson aggregated all word classes across all age groups. Thus, Dalton and Richardson's core lexicon lists can be applied to PWA of all ages. However, since Kim and colleagues provide separate core lexicon lists by age, SLPs should make sure to use core lexicon lists applicable to the patient's respective age group. Researchers investigating core lexicon measures have clarified the common, simple rules for scoring PWA's word retrieval performance in discourse, which follows:

- Synonyms are not counted, due to the importance of producing the specific target words.
- Plurals, verb conjugations, and inflections for the target core lexicon are scored.
- Only one point is given, regardless of frequency of a target word presented in a language sample.

Task instructions provided to PWA are distinct depending on discourse elicitation tasks. Table 1 provides detailed instructions that researchers have used in their studies, which SLPs can apply for clinical practices. Instructions for Broken Window and the Cinderella story followed the AphasiaBank protocols. See Appendices in the study by Dalton et al for core lexicon checklists reported by previous studies.[66]

Considerations for Assessment

In the clinical setting, SLPs may be confronted with referrals for patients who report language difficulties, but performance on standardized measures does not demonstrate language impairment. For example, a patient may report language difficulties suggesting presence of aphasia, but scores are at or above the diagnostic cutoff on the WAB-R. Fromm et al recently shed light on the importance of capturing subtle language deficits in these patients, indicating that standardized test batteries are not sensitive to subtle linguistic deficits that can affect discourse-level language production.[67] Returning back to Dalton and Richardson's study,[54] omnibus median test demonstrated that core lexicon measures differed the control group from all aphasia subtypes, including those who have had a stroke but are not aphasic by WAB (NABW; $N = 25$). The core lexicon measure also differed between the NABW and persons with anomic ($p = 0.002$), conduction ($p < 0.001$), and Wernicke's aphasia ($p < 0.001$) groups. The researchers did not intend to provide such clear-cut classification of aphasia types based on the number of core lexicon items produced by speakers. However, this suggests that core lexicon measures appear well suited for measuring subtle communicative deficits in such patients.

Another practical implication of using core lexicon measures within clinical settings is that core verb lists can provide a sensitive account of

Table 1 Core Lexicon Protocols

Tasks	Authors	Instructions
Broken Window[a]	Dalton and Richardson[54]	"Now I'm going to show you these pictures." (Present picture series) "Take a little time to look at these pictures. They tell a story. Take a look at all of them, and then I'll ask you to tell me the story with a beginning, a middle, and an end. You can look at the pictures as you tell the story."
Cinderella[a]	Hudspeth et al[24]	"I am going to ask you to tell a story. Have you ever heard the story of Cinderella?" (Make note of answer for demographic data. If answer is no, ask participant to tell a fairy tale s/he knows.) "Do you remember much about it? These pictures might remind you of how it goes. Take a look at the pictures and then I'll put the book away, and ask you to tell me the story in your own words." Allow participant to look through book (assist with page turning if needed) and then, if necessary, prompt: "Now tell me as much of the story of Cinderella as you can. You can use any details you know about the story, as well as the pictures you just looked at." If participant gives a response of fewer than three utterances, or seems to falter, allow 10 seconds, then prompt: "What happened next?" or "Go on." Continue until participant concludes story or it is clear s/he has finished. If no response, go to Troubleshooting questions.
Wordless Picture books (Good Dog Carl and Picnic)	Kim et al[25]	"These are children's books without words so that a person can make up their own story. First, you will look through the book and get an idea of the story. Then, you will start at the beginning and tell me the story that goes with the pictures." "Look at this book, and when you are ready, tell me the story that goes with the pictures."

[a]AphasiaBank protocols.

overall language severity. Verbs are viewed to be the building blocks and/or central themes of utterances. Considering that verb representations are affected by knowledge of both lexical and syntactic information, verbs should be considered at the utterance level for assessment as well as treatment.[68] In Kim et al,[29] they conducted multiple stepwise regression analyses to investigate which variables predicted overall aphasia severity, as determined by WAB-R AQs, among the five core lexicon items (verbs, nouns, adjectives, adverbs, and function words) for two stories (*GDC*, *Picnic*). For both stories, results indicated that verbs were a significant predictor of WAB-R AQs. For *GDC*, core verbs alone explained 70% of the variance in overall aphasia severity. For *Picnic*, core verbs explained 81% of the variance. This is not to say that core verb list should replace the language battery measure, but it can be completed in shorter time than the WAB, providing potentially general information on language performance.

Subsequently, Kim et al extended the scope of use of core lexicon measures by applying the measures developed based on language samples obtained from wordless pictures books (*GDC*, *Picnic*) to the language samples from the Cinderella story.[69] Because a limited number of function words are used in our daily life,[70,71] the researchers hypothesized that core function word lists could be applied to generic language samples, regardless of elicitation task and patient age. Moreover, simplified utterances with the omission of function words is a typical feature of nonfluent types of aphasia.[72,73] For these reasons, the researchers attempted to investigate diagnostic accuracy for fluency. A total of 208 PWA Cinderella story samples (fluent, 110; nonfluent, 98) were retrieved from AphasiaBank. Identification accuracy of fluency was investigated using receiver

Table 2 Demographic Data from AphasiaBank

Group	*N* (F:M)	Age (SD)	Education (SD)	Function words (SD) (%)
Fluent	110 (49:61)	62.0 (16.5)	15.5 (3.0)	62 (16.5)
Nonfluent	98 (46:52)	57.6 (12.6)	15.3 (2.5)	36.9 (21.8)
Mean	208	60.8	15.4	50.2
SD	(95:113)	13.0	2.8	22.9

operating characteristic (ROC) curve analysis.[74] A cut-off score of 12 (out of 25) produced a sensitivity of 82.7% and specificity of 65.3%. The area under the ROC curve (AUC)[75] was 0.814 (95% CI:0.757, 0.871, SE = 0.029, $p < 0.001$), suggesting that the core set of function words was an accurate classifier for differentiating participants with fluent aphasia from participants with nonfluent aphasia. Using the core lexicon list, 82.7% of those who are nonfluent aphasia were correctly identified as being nonfluent aphasia. Of all PWA with fluent type of aphasia, 65.3% were correctly identified as not being nonfluent aphasia. Using the cut-off score of 12 core function word items, the identification accuracy was 84.1%. These findings demonstrate that core function word lists may be more helpful when a quick screening tool is necessary to identify fluency in PWA (Tables 2 and 3).

In addition to the diagnostic purpose of core lexicon lists, we hypothesize that the effectiveness of treatment may be measured using core lexicon measures. There is general agreement in the literature that generalization effects of treatment should be measured at the discourse level, even when interventions are implemented at the word level.[76–80] In most cases, PWA and their families' ultimate goal of treatment is to improve their ability to communicate with others in socially framed situations, which has immediate relevance to discourse-level language performance.[81] In doing so, SLPs may also predict PWA's communicative ability in a less controlled context through discourse-level assessments following the treatment.[76] For example, semantic treatments, such as Semantic Feature Analysis (SFA) treatment for nouns and for verbs, have been frequently used to improve semantic networks, which in turn enhances PWA's ability to retrieve words.[82] The degree to which such semantic treatment at single-word or sentence levels generalize to discourse-level production may be a clinically and personally vital question for SLPs, PWA, and caregivers. Conceptually, core lexicon measures are considered to tap into lexical semantics. When producing a target word (target lexical item), the features for the concept of the target words are activated, and the activation spreads to the item in the lexicon. Of all the lexicon items being activated, a lexical item that receives the greatest activation is selected. The activation from the selected item propagates to the phonological representation to produce the target word. Following this conceptualization, it seems reasonable that core lexicon measures are particularly appropriate for use in examining generalized improvement to discourse.

Table 3 Performance Measures of Receiver Operating Characteristic Curve

	Value	Lower limit	Upper limit
Sensitivity	0.827	0.743	0.893
Specificity	0.653	0.550	0.746
Positive predictive value	0.728	0.635	0.823
Negative predictive value	0.771	0.761	0.840

Unlike nouns and verbs that are regarded as the major class of content words, modifiers have received considerably less attention in research and clinical practices. Penn suggested that an increased use of adjectives reflects elaboration of verbal messages produced by PWA.[83] In a later treatment study, an extensive array of rehabilitation services (e.g., linguistic deficit-based therapy, conversational practice, and training in coping strategies) was offered to 18 PWA who were tested at 3-month intervals for a year. A significantly greater number and proportion of modifiers (adjectives and adverbs) were found at the end of treatment compared with pretreatment. The researchers concluded that

production of modifiers manifested qualitative changes in PWA's language gain over the course of language treatment. Additionally, in studies of second language learning, it has been suggested that adverbs serve as an integral device to measure lexical variation[84] and language proficiency.[85] Though studies of modifier production in aphasia are scant, and core adjective lists have not been used to document treatment effectiveness, it still may be beneficial to examine their quantitative differences following the treatment sessions.

FUTURE DIRECTIONS AND CONCLUSIONS

Previous theoretically developed discourse outcome measures have failed to meaningfully impact clinical settings. Unlike theoretically focused measures, core lexicon measures are empirically driven procedures, based on the principle that discourse disruptions featured in PWA lie on the continuum of normal discourse. Katz et al reported that the average number of initial assessment sessions implemented are reported to be one or two for inpatients with acute aphasia and outpatients with chronic aphasia in the U.S. Veteran Affairs and U.S. private sectors.[48] In the short time available for assessment, core lexicon measures may serve as an alternative, ecological approach to the assessment of language impairments at the discourse level for SLPs requiring a clinician-friendly discourse measure.

Core lexicon measures are a relatively new method and as such, few studies exist that have used the measure to quantify discourse ability in PWA. Various discourse tasks and different criteria to select lexical items to be included in the measure have been used. The next logical step, then, is to investigate these factors which might lead to different outcomes. Currently, it is unknown which criterion is better for constructing core lexicon measures to accurately quantify word retrieval ability at the discourse level. Furthermore, PWA's discourse performance differs across different discourse elicitation tasks that vary in cognitive and linguistic demands on speakers.[86,87] The degree of contextual support provided in the illustrations may make it possible that PWA retrieve core lexicon items more easily for one elicitation task compared with another. Finally, for core lexicon measures to be used in everyday clinical settings, it is imperative to examine the psychometric properties of the core lexicon items, which will help achieve more effective, precise outcomes with greater measurement precision.

DISCLOSURES

H.K. receives a graduate assistantship. She has no other financial and nonfinancial disclosures. H.H.W. receives a salary from East Carolina University. She has no other financial and nonfinancial disclosures.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Brady MC, Godwin J, Enderby P, Kelly H, Campbell P. Speech and language therapy for aphasia after stroke: an updated systematic review and meta-analyses. Stroke 2016;47(10):e236–e237
2. Dietz A, Boyle M. Discourse measurement in aphasia research: have we reached the tipping point? Aphasiology 2018;32(04):459–464
3. 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
4. Maddy KM, Howell DM, Capilouto GJ. Current practices regarding discourse analysis and treatment following non-aphasic brain injury: a qualitative study. J Interact Res Commun Disord 2015;6(02):211
5. Bryant L, Spencer E, Ferguson A. Clinical use of linguistic discourse analysis for the assessment of language in aphasia. Aphasiology 2017;31(10): 1105–1126
6. Coelho CA. Management of discourse deficits following traumatic brain injury: progress, caveats, and needs. Semin Speech Lang 2007;28(02): 122–135
7. Kintz S, Wright HH. Discourse measurement in aphasia research. Aphasiology 2018;32(04):472–474
8. Wallace SJ, Worrall L, Rose T, Le Dorze G. Measuring outcomes in aphasia research: A review of current practice and an agenda for standardisation. Aphasiology 2014;28(11):1364–1384
9. de Riesthal M, Diehl SK. Conceptual, methodological, and clinical considerations for a core outcome set for discourse. Aphasiology 2018;32(04): 469–471

10. Dietz A, Boyle M. Discourse measurement in aphasia: consensus and caveats. Aphasiology 2018;32(04):487–492
11. Wallace SJ, Worrall LE, Rose T, Le Dorze G. Discourse measurement in aphasia research: have we reached the tipping point? A core outcome set… or greater standardisation of discourse measures?. Aphasiology 2018;32(04):479–482
12. Kurland J, Stokes P. Let's talk real talk: an argument to include conversation in a D-COS for aphasia research with an acknowledgment of the challenges ahead. Aphasiology 2018;32(04):475–478
13. Armstrong L, Brady M, Mackenzie C, Norrie J. Transcription-less analysis of aphasic discourse: a clinician's dream or a possibility? Aphasiology 2007;21(3–4):355–374
14. Boles L, Bombard T. Conversational discourse analysis: appropriate and useful sample sizes. Aphasiology 1998;12(7–8):547–560
15. Elia D, Liles BZ, Duffy RJ, Coelho CA, Belanger SA. An investigation of sample size in conversational analysis. I Paper presented at: American Speech-Language-Hearing Association; November 1994; New Orleans, LA
16. Loban W. Language Development: Kindergarten through Grade Twelve. Urbana, IL: National Council of Teachers of English; 1976
17. Hunt KW. Sentence Structures Used by Superior Students in Grades Four and Twelve, and by Superior Adults. Tallahassee, FL: Florida State University; 1966
18. Kroll B. Ways communicators encode propositions in spoken and written English: a look at subordination and coordination. In: Discourse Across Time and Space. Los Angeles, CA: University of Southern California; 1977:69–108
19. Chafe W. Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing. Chicago, IL: University of Chicago Press; 1994
20. Halliday MAK. Intonation and Grammar in British English. The Hague, Paris: Mouton; 1970
21. 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
22. MacWhinney B, Fromm D, Holland A, Forbes M, Wright H. Automated analysis of the Cinderella story. Aphasiology 2010;24(6-8):856–868
23. Fromm DA, Forbes M, Holland A, MacWhinney B. PWAs and PBJs: Language for describing a simple procedure. Paper presented at: Clinical Aphasiology Conference; May 2013; Tucson, AZ
24. Hudspeth SG, Dillow E, Richardson JD. Core lexicon analysis: efficient assessment of narrative discourse in persons with aphasia. Paper presented at: American Speech-Language-Hearing Association; November 2013; Chicago, IL
25. 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
26. Day A. Good Dog, Carl. New York: Scholastic; 1985
27. McCully EA. Picnic. New York: Harper & Row; 1984
28. Kertesz A. Western Aphasia Battery–Revised (WAB-R). San Antonio, TX: Pearson; 2006
29. Kim H, Kintz S, Wright HH. Function words in narrative discourse in aphasia. Paper presented at: Clinical Aphasiology Conference. May 2017; Salt Lake, UT
30. Cui G, Zhong X. Adaptation in aphasia: revisiting language evidence. Aphasiology 2018;32(08): 855–875
31. Dalton SGH, Hubbard HI, Richardson J. Moving towards non-transcription based discourse analysis in stable and progressive aphasia. Semin Speech Lang 2020;41:32–44
32. Chen S, Bates E. The dissociation between nouns and verbs in Broca's and Wernicke's aphasia: findings from Chinese. Aphasiology 1998;12(01):5–36
33. Caramazza A, Hillis AE. Lexical organization of nouns and verbs in the brain. Nature 1991;349 (6312):788–790
34. Bates E, Chen S, Tzeng O, Li P, Opie M. The noun-verb problem in Chinese aphasia. Brain Lang 1991;41(02):203–233
35. Kim M, Thompson CK. Patterns of comprehension and production of nouns and verbs in agrammatism: implications for lexical organization. Brain Lang 2000;74(01):1–25
36. Luzzatti C, Chierchia G. On the nature of selective deficits involving nouns and verbs. Ital J Linguist 2002;14:43–72
37. Berndt RS, Haendiges AN, Mitchum CC, Sandson J. Verb retrieval in aphasia. 2. Relationship to sentence processing. Brain Lang 1997;56(01): 107–137
38. Zingeser LB, Berndt RS. Grammatical class and context effects in a case of pre-anomia: implications for models of language processing. Cogn Neuropsychol 1988;5:473–516
39. Saffran EM, Schwartz MF, Marin OSM. Evidence from aphasia: isolating the components of a production model. Lang Prod. 1980;1:221–241
40. Milman L, Clendenen D, Vega-Mendoza M. Production and integrated training of adjectives in three individuals with nonfluent aphasia. Aphasiology 2014;28(10):1198–1222
41. Sarno MT, Postman WA, Cho YS, Norman RG. Evolution of phonemic word fluency performance in post-stroke aphasia. J Commun Disord 2005;38 (02):83–107
42. Meltzer-Asscher A, Thompson CK. The forgotten grammatical category: Adjective use in agrammatic aphasia. J Neurolinguist 2014;30:48–68

43. Macoir J, Laforce R Jr, Brisson M, Wilson MA. Preservation of lexical-semantic knowledge of adjectives in the semantic variant of primary progressive aphasia: Implications for theoretical models of semantic memory. J Neurolinguist 2015;34:1–14
44. Kaplan E, Goodglass H, Weintraub S. Boston Naming Test. Philadelphia, PA: Pro-Ed.; 2001
45. Guo YE, Togher L, Power E. Speech pathology services for people with aphasia: What is the current practice in Singapore? Disabil Rehabil 2014;36(08):691–704
46. Verna A, Davidson B, Rose T. Speech-language pathology services for people with aphasia: A survey of current practice in Australia. Int J Speech Lang Pathol 2009;11(03):191–205
47. Beukelman D, Mirenda P. Augmentation and alternative communication: Management of severe communication disorders in children and adults. Baltimore, MD: Brooks; 1998
48. Katz RC, Hallowell B, Code C, et al. A multinational comparison of aphasia management practices. Int J Lang Commun Disord 2000;35(02): 303–314
49. Ivanova MV, Hallowell B. A tutorial on aphasia test development in any language: Key substantive and psychometric considerations. Aphasiology 2013;27(08):891–920
50. Cronbach LJ, Furby L. How we should measure" change": Or should we? Psychol Bull 1970;74(01): 68
51. Nunnally JC, Bernstein IH. Psychometric Theory (McGraw-Hill Series in Psychology). Vol 3. New York: McGraw-Hill; 1994
52. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(02): 420–428
53. Kimberlin CL, Winterstein AG. Validity and reliability of measurement instruments used in research. Am J Health Syst Pharm 2008;65(23): 2276–2284
54. 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
55. Kim H, Wright HH. Concurrent validity and reliability of the Core Lexicon measure as a measure of word retrieval ability in aphasia narratives. Am J Speech Lang Pathol 2019;(5):1–10
56. Dijkstra K, Bourgeois MS, Allen RS, Burgio LD. Conversational coherence: discourse analysis of older adults with and without dementia. J Neurolinguist 2004;17(04):263–283
57. 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
58. Schneider P, Dubé RV, Hayward D. The Edmonton Narrative Norms Instrument. Retrieved from University of Alberta Faculty of Rehabilitation Medicine. Available at: http:://rehabmed.ualberta.ca/spa/enni2005. Accessed November 14, 2019
59. Wright HH, Capilouto GJ. Considering a multilevel approach to understanding maintenance of global coherence in adults with aphasia. Aphasiology 2012;26(05):656–672
60. Covington MA. MATTR User Manual. Athens, Georgia: University of Georgia Artificial Intelligence Center; 2007
61. Covington MA, McFall JD. Cutting the Gordian knot: the moving-average type–token ratio (MATTR). J Quant Linguist 2010;17(02):94–100
62. Fergadiotis G, Wright HH. Lexical diversity for adults with and without aphasia across discourse elicitation tasks. Aphasiology 2011;25(11):1414–1430
63. Glosser G, Deser T. A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. J Gerontol 1992;47(04):266–272
64. Marini A, Boewe A, Caltagirone C, Carlomagno S. Age-related differences in the production of textual descriptions. J Psycholinguist Res 2005;34(05): 439–463
65. Kintz S, Hibbs V, Henderson A, Andrews M, Wright HH. Discourse-based treatment in mild traumatic brain injury. J Commun Disord 2018; 76:47–59
66. Dalton SGH, Kim H, Richardson J, Wright H. A compendium of core lexicon checklists. Semin Speech Lang 2020;41:45–60
67. 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
68. Conroy P, Sage K, Lambon Ralph MA. Towards theory-driven therapies for aphasic verb impairments: a review of current theory and practice. Aphasiology 2006;20(12):1159–1185
69. Kim H, Kintz S, Wright HH. Development of a core function word set for clinical use. Paper presented at: Clinical Aphasiology Conference; Whitefish, MT; 2019
70. Baayen R, Piepenbrock R, Gulikers L. The CELEX lexical database (Release 2). Linguist Data Consortium. Philadelphia, PA: University of Pennsylvania; 1995
71. Chung C, Pennebaker J. The psychological functions of function words. Soc Commun. 2011;1:343–359
72. Gordon JK. Measuring the lexical semantics of picture description in aphasia. Aphasiology 2008; 22(7-8):839–852
73. Manning M, Franklin S. Cognitive grammar and aphasic discourse. Clin Linguist Phon 2016;30(06): 417–432

74. Macmillan NA, Creelman CD. Detection Theory: A User's Guide. New York: Cambridge University Press; 1991
75. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143(01):29–36
76. Conroy P, Sage K, Ralph ML. Improved vocabulary production after naming therapy in aphasia: can gains in picture naming generalize to connected speech? Int J Lang Commun Disord 2009;44(06):1036–1062
77. Law SP, Kong APH, Lai LWS, Lai C. Effects of context and word class on lexical retrieval in Chinese speakers with anomic aphasia. Aphasiology 2015;29(01):81–100
78. Ferguson A, Armstrong E. The Palpa: a valid investigation of language? Aphasiology 1996;10 (02):193–197
79. Mayer J, Murray L. Functional measures of naming in aphasia: word retrieval in confrontation naming versus connected speech. Aphasiology 2003;17(05): 481–497
80. Pashek GV, Tompkins CA. Context and word class influences on lexical retrieval in aphasia. Aphasiology 2002;16(03):261–286
81. Cruice M, Worrall L, Hickson L, Murison R. Finding a focus for quality of life with aphasia: social and emotional health, and psychological well-being. Aphasiology 2003;17(04):333–353
82. Boyle M, Coelho CA. Application of semantic feature analysis as a treatment for aphasic dysnomia. Am J Speech Lang Pathol 1995;4(04): 94–98
83. Penn C. Compensation and language recovery in the chronic aphasic patient. Aphasiology 1987;1 (03):235–245
84. Lu X. The relationship of lexical richness to the quality of ESL learners' oral narratives. Mod Lang J 2012;96(02):190–208
85. Grant L, Ginther A. Using computer-tagged linguistic features to describe L2 writing differences. J Second Lang Writ 2000;9(02):123–145
86. Armstrong E. Aphasic discourse analysis: the story so far. Aphasiology 2000;14(09):875–892
87. Coelho CA. Story narratives of adults with closed head injury and non-brain-injured adults: influence of socioeconomic status, elicitation task, and executive functioning. J Speech Lang Hear Res 2002;45 (06):1232–1248