Core lexicon in aphasia: A longitudinal study

Hana Kim, Shauna Berube & Argye E. Hillis

Routledge
Taylor & Francis Group

Check for updates

Core lexicon in aphasia: A longitudinal study

Hana Kim [a,b], Shauna Berube[b] and Argye E. Hillis [b,c,d]

aDepartment of Communication Sciences and Disorders, University of South Florida, Tampa, FL 33620;
bDepartment of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21287;
cDepartment of Physical Medicine and Rehabilitation, Johns Hopkins University School of Medicine,
Baltimore, MD 21287; dDepartment of Cognitive Science, Krieger School of Arts and Sciences, Johns Hopkins
University, Baltimore, MD 21218

ABSTRACT

Background: General consensus exists between clinicians as to the incorporation of discourse outcome measures into language assessment for persons with aphasia (PWA). The development of core lexicon measures (CoreLex) has enabled clinicians to reduce time and labor intensive preparatory work for discourse analysis, which has been considered as an alternative measure to quantify word retrieval ability in discourse in a clinical context. Although previous studies have investigated the quality of the measure, CoreLex has rarely been longitudinally explored.

Aims: We aimed to investigate the adequacy of CoreLex to document linguistic changes in PWA over time. Specifically, we examined (1) whether natural language recovery from acute to chronic stages is manifested differentially by tasks and (2) the extent to which the ability to retrieve words in isolation predicts the ability to retrieve words in context.

Methods: A total of 19 PWA participated in the study. They completed a language assessment including confrontation naming tasks (Boston Naming Test [BNT] and Hopkins Action Naming Assessment [HANA]) and a picture description task using the Cookie Theft picture at acute and chronic stages. Discourse samples from the picture description task were quantified using CoreLex.

Results: We found significant differences across tasks and timepoints by PWA. Moderate correlations between the confrontation naming tasks and CoreLex were found at the acute stage but not at the chronic stage. Additionally, McNemar's tests demonstrated a significant difference in PWA's performance in CoreLex from the acute to the chronic stages.

Conclusions: Our findings show that performance by PWA improves over time on all tasks, but language gains are manifested differentially by tasks. Performance in confrontation naming moderately predicts word retrieval in context acutely. However, lack of correlations between confrontation naming tasks and CoreLex later endorse inadequacy of using confrontation naming tasks as a proxy measure for discourse-level performance and improvement for PWA.

CONTACT Hana Kim ✉ hanakim1@usf.edu 🖳 Department of Communication Sciences and Disorders, 4202 East Fowler Avenue, PCD 1017, Tampa, FL 33620

Introduction

The primary goal of speech-language therapy for persons with aphasia (PWA) is to evaluate communication abilities, and provide treatment based on the evaluation to help patients build communicative competence (Simmons-Mackie & Damico, 2010). Successful service delivery also entails adjusting therapy goals with frequent assessment (Alary Gauvreau et al., 2019). Practically, confrontation naming tasks have frequently been used as primary outcome measures in clinical settings, with clinicians tending to rely on results from them to draw clinical decisions (Fergadiotis & Wright, 2016). Although confrontation naming tasks have many advantages, such as high test-rest reliability and simplicity to administer and interpret the results (Fergadiotis et al., 2015, 2019; Herbert et al., 2008), criticisms are directed toward their lack of ecological validity.

Evidence shows that the nature of the relationship between performance on confrontation naming tasks and discourse performance is unclear. A recent review article has reported mixed correlational results between confrontation naming performance and discourse performance by measures (Mason & Nickels, 2022). For example, the authors identified 39 different discourse measures used across studies, and a strong, reliable correlation across discourse outcomes is lacking. Pashek and colleagues (2002) reported that the regression coefficient between confrontation naming accuracy on the Boston Naming Test (Kaplan et al., 2001) and the proportion of word finding difficulty in discourse was modest ($r^2 = 0.58$). Mayer and Murray (2003) examined the relationship using the proportion of substantive and light verbs in discourse, and the range of correlations (r^2) varied between 0.46 and 0.79. Fergadiotis et al. (2019) estimated discourse performance by using the percentage of correct information units (CIUs). They found that naming ability at the single word level accounted for 63% of the variance in discourse informativeness. Therefore, reliance on confrontation naming tasks to predict discourse-level word retrieval ability may be a flawed approach.

Moreover, cross-sectional research has dominated previous investigations on the relationship of word retrieval between isolation and context. In general, PWA experience some recovery in language function following stroke. However, the degree of recovery is affected by various factors, such as demographic variables and lesion-related factors (e.g., lesion volume, lesion location), and is highly variable from patient to patient. As such, it is critical for clinicians to understand their clients' current status of language recovery prior to making any clinical decisions. To the best of our knowledge, only one study has longitudinally reported comparisons between scores on confrontation naming tasks and discourse tasks in 10 PWA (Herbert et al., 2008). Naming performance was assessed using a set of 200 black and white line drawing pictures. Discourse samples were naturally elicited without pre-determined topics, and the PWA and their conversation partners were instructed to converse as normally as possible. Discourse measures included speech units, the proportion of turn taking, and the number of content words. They found significant correlations between picture-naming scores and five variables of the discourse measures at the first assessment, and a reduced number of correlations were found at the second assessment. The range of correlations (r^2) varied between 0.55 to 0.72. This study provided longitudinal investigations of a discourse measure, but the ultimate focus of the study was to establish the reliability and stability of the measure. The time from the

first assessment to the second assessment was approximately 8 days, and both assessments were administered at the chronic stage.

Despite the rich history of discourse measures in research, there are few developed discourse outcome measures considering clinical usability. Core lexicon measure (CoreLex) is one discourse measure which was intended to provide a clinician-friendly means to quantify word retrieval ability in discourse based on normal expectations of discourse production for specific discourse elicitation tasks (Dalton et al., 2020; Kim et al., 2019). Importantly, scoring for CoreLex is relatively simple in that it accompanies a checklist of lexical items which assesses whether the specific lexical item was present or not in discourse samples. During the process of scoring, subjective judgements to determine acceptable words/synonyms corresponding to discourse stimuli are not required, which resembles the use of confrontation naming tasks. CoreLex is also reliable and sensitive to measuring overall language severity. Inter-rater reliability and concurrent validity with other existing discourse measures have been demonstrated (Kim & Wright, 2020a; Kim et al., 2022). Previous research has demonstrated that CoreLex captures aphasia severity (Kim et al., 2019) and differentiates PWA from cognitively healthy controls and subtypes of aphasia (Dalton & Richardson, 2015).

Although previous studies have shown validity and reliability in CoreLex, its responsiveness to natural recovery in language is yet to be explored. Further, it is important to examine the association between CoreLex and confrontation naming tasks over time. This would inform whether naming tasks can be an appropriate surrogate measure for predicting discourse-level word retrieval ability in PWA. Accordingly, the aims of this investigation were to investigate: 1) whether natural language recovery from acute to chronic stages is manifested differentially by tasks (two confrontation naming tasks and CoreLex), and 2) the extent to which confrontation naming test scores correlate with performance of word retrieval in discourse captured by CoreLex at acute and chronic stages.

Methods

Participants

A total of 19 participants with aphasia were included in the current study. They represent a subset of a larger study directed by the third author, which sought to investigate the longitudinal recovery of aphasia. We included only those who completed the language testing both at the acute stage (mean months post-stroke onset = 0.11) and chronic stage (mean months post-stroke onset = 8.35). Participants in the acute phase of recovery were defined as those who completed language testing within the first week following a stroke. Participants in the chronic phase of recovery were tested at least 6 months after the occurrence of a stroke. All participants were right-handed, native English speakers. Anyone who had a history of co-occurring neurological diagnoses affecting the brain, vision and/or hearing problems were excluded. All work was conducted in compliance with the approval of Johns Hopkins School of Medicine Institutional Review Board. Demographic information can be found in Table 1.

Table 1. Demographic information

Participant ID	Age (in years)	Gender	Education (in years)	Mean months post onset (T1)	Mean months post onset (T2)	AQ (T1)	BDAE scale (T1)	Aphasia Type (T1)
A01	83	F	12	0.07	5.60	90.9		Anomic
A02	63	F	16	0.00	5.20	86		Anomic
A03	80	F	11	0.30	5.93	89		Anomic
A04	74	F	14	0.07	6.53	77.3		Transcortical Motor
A05	50	F	10	0.33	5.83	10		Broca
A06	56	M	14	0.07	5.23		1	Global
A07	62	M	14	0.07	7.70	7.5		Global
A08	70	F	12	0.17	11.33	54.6		Broca
A09	79	F	17	0.20	5.80	78.7		Conduction
A10	87	M	20	0.07	6.17		4	Conduction
A11	68	M	18	0.17	7.07	90		Anomic
A12	57	M	14	0.30	14.37	46.2		Broca
A13	58	M	12	0.07	6.37		1	Global
A14	77	F	11	0.00	5.40		5	Anomic
A15	64	M	10	0.13	4.17		4	Anomic
A16	47	M	15	0.03	15.97		5	Anomic
A17	61	F	9	0.07	14.77		4	Anomic
A18	85	F	12	0.03	12.37	88.8		Broca
A19	66	M	9	0.03	12.93		5	Anomic

Note. AQ = Aphasia Quotient from Western Aphasia Battery (Kertesz, 2006); BDAE = Boston Diagnostic Aphasia Examination (Goodglass et al., 2001); T1 = Acute stage; T2 = Chronic stage

Language assessments

Single word level (Confrontation naming tasks)

We used noun and verb naming tasks for measuring single word level performance. For the noun naming task, we used the Boston Naming Test (BNT; Kaplan et al., 2001) because of its frequent use in clinics. For the verb naming task, the Hopkins Action Naming Assessment (HANA; Breining et al., 2015) was used because it captures verb naming which can be selectively impaired after stroke (e.g., Hillis, 2007). The BNT is one of the most consistently used standardized measures of language performance in clinical settings (Kiran et al., 2018), focusing on noun naming ability. It provides black and white line drawings of 30 objects in a short version. HANA has a similar method of delivery as the BNT in that it consists of black and white images of 30 items that match in frequency and length to the short version of the BNT using a psycholinguistic database (e.g., N-Watch, CELEX English linguistic database) (Breining et al., 2021). Breining and colleagues (2021) demonstrated the two assessments do not statistically differ on the length and frequency of the items. Although HANA is a comparatively new tool to detect verb naming difficulties, it has been utilized in conjunction with other assessment tools in research (e.g., Keator et al., 2020; Long et al., 2018), and its clinical usability has recently been demonstrated (Breining et al., 2021). In the current study, scores for the BNT and HANA for each participant were transformed into the proportion of accurate responses, and then used for statistical analysis.

Word retrieval ability at discourse levels (Core lexicon measure)

The concept of core lexicon was first introduced in aphasia research to quantify word retrieval ability in spoken discourse (MacWhinney et al., 2010). CoreLex provides a checklist that consists of critical lexical items required to deliver a story in a coherent

way. The "core lexical items" are selected based on performance in normative samples. Since the measurement involves checking off any core lexicon items produced in patients' language samples, and each item is given one point, it has been considered to have potential to be widely used in clinical settings (Dalton et al., 2020b; Kim & Wright, 2020b). In the current study, we used a CoreLex checklist developed by Dalton and colleagues (Dalton et al., 2021) for the Cookie Theft picture from the Boston Diagnostic Aphasia Examination (Goodglass et al., 2001).

Statistical Analyses

A two-factor within-subject analysis of variance (ANOVA) was computed. Factors were time (acute and chronic) and task (BNT, HANA, and CoreLex). A secondary analysis was conducted to examine a group-level changes in individual performance using McNemar's test. To perform this test, each score was converted into a binary rating (0 or 1) based on the mean and standard deviation for each task. Scores below two standard deviations from the control mean are generally considered abnormal performance in language assessments, which were coded 0. Scores above the threshold of -2.0 standard deviations were coded 1. Pearson correlation coefficients were computed to examine the relationship between performance from confrontation naming tasks and performance from a discourse task at the acute and chronic stages, respectively. Analyses were computed using SPSS version 27.

Results

Analysis of variance

Main effect for the task was significant, $F(2, 36) = 5.49$, $p < .01$. Paired t-test with Bonferroni correction revealed significant differences between BNT and HANA ($p < 0.05$). PWA performed more accurately on BNT (mean = 65.35% correct) than on HANA (mean = 59.90% correct). No difference was found between the confrontation naming tasks and CoreLex. Main effect for time was significant, $F(1, 18) = 28.9$, $p < 0.001$. Performance at the chronic stage was significantly better than performance at the acute stage across all three tasks (46.48% vs 68.54% correct). A two-way interaction between the task and time was not statistically significant, $F(2, 36) = 2.05$, $p = 0.14$. See Table 2.

McNemar's test

McNemar's test was additionally conducted to examine group-level changes in PWA's performance between the acute and chronic stages by task. A significant change for CoreLex was found between the two time points, χ^2 (1, $N = 19$) = 8.1, $p = .002$. This indicates that there was a considerable change in PWA's performance captured by CoreLex from acute to chronic stages. Based on CoreLex performance, 4 PWA (21.1%) showed normal performance, as they were no more than 2 SD below average at the acute stage, while 14 PWA (73.7%) showed normal performance at the chronic stage. No significant changes between the two time points were found for BNT, $\chi^2(1, N = 19)$ = 1.33, $p = .250$, or HANA, χ^2 (1, $N = 19$) = 2.25, $p = .125$.

Table 2. Performance of persons with aphasia by task

Participant ID	BNT (T1)	HANA (T1)	CoreLex (T1)	BNT (T2)	HANA (T2)	CoreLex (T2)
A01	80.0	56.7	42.3	100.0	93.3	65.4
A02	90.0	6.7	53.9	93.3	90.0	65.4
A03	70.0	50.0	26.9	70.0	47.0	23.1
A04	73.3	76.7	53.8	76.7	83.3	73.1
A05	0.0	2.8	46.2	70.0	79.3	80.8
A06	6.7	0.0	19.2	43.3	28.6	50.0
A07	3.3	0.0	11.5	47.0	47.0	76.9
A08	70.0	36.7	11.5	93.0	73.0	53.8
A09	86.7	73.3	50.0	86.7	73.3	65.4
A10	66.7	60.0	57.7	90.0	80.0	65.4
A11	90.0	76.7	76.9	97.0	90.0	80.8
A12	10.0	13.3	7.7	53.0	53.0	73.1
A13	30.0	22.9	0.0	50.0	53.3	23.1
A14	90.0	71.0	42.3	86.7	83.3	50.0
A15	76.6	54.3	30.8	76.6	74.3	57.7
A16	90.0	83.0	53.8	93.0	83.0	57.7
A17	50.0	65.7	19.2	53.3	62.9	57.7
A18	66.7	60.0	65.4	80.0	66.7	73.1
A19	33.3	25.7	61.5	43.3	65.7	84.6
Mean (SD)	57.0 (32.6)	44.0 (29.1)	38.5 (22.2)	73.8 (19.7)	69.8 (17.5)	61.9 (17.1)

Note. Raw scores were converted into percentages. BNT = Boston Naming Test (Kaplan et al., 2001); HANA = Hopkins Action Naming (B L Breining et al., 2015); T1 = acute stag; T2 = chronic stage

Correlation coefficients

At the acute stage, significant correlations were found between BNT and HANA, $r = 0.79$, $p < .001$, BNT and CoreLex, $r = 0.53$, $p < .05$, and HANA and CoreLex, $r = 0.49$, $p < .05$. This suggests positive relationships among the three measures (CoreLex, BNT, HANA) at the acute stage. At the chronic stage, significant correlations were only found between BNT and HANA, $r = .826$, $p < .001$, and no other significant correlations were found (See Table 3). This indicates a positive association only between BNT and HANA at the chronic stage.

Individual analyses

Raw scores of individual participants show that five out of six PWA who demonstrated either a decline or no change in either BNT or HANA showed an increase in CoreLex at T2 relative (See Figures 1, 2, & 3). Out of these five PWA, four had anomic aphasia based on their language test results at T1. The other PWA who was categorized as conduction aphasia showed no changes in either the BNT or HANA from T1 to T2, but improvement in performance was observed in CoreLex from T1 (50%) to T2 (65%).

Table 3. Correlations between BNT, HANA, and CoreLex

	BNT (T1)	HANA (T1)	CoreLex (T1)	BNT (T2)	HANA (T2)	CoreLex (T2)
BNT (T1)		.79**	.53*	.83**	.69**	-.12
HANA (T1)			-.49*	.59**	.53*	-.05
CoreLex (T1)				.55*	.66**	.54*
BNT (T2)					.83**	.09
HANA (T2)						.38

Note. BNT = Boston Naming Test (Kaplan et al., 2001); HANA = Hopkins Action Naming (B L Breining et al., 2015); T1 = acute stag; T2 = chronic stage
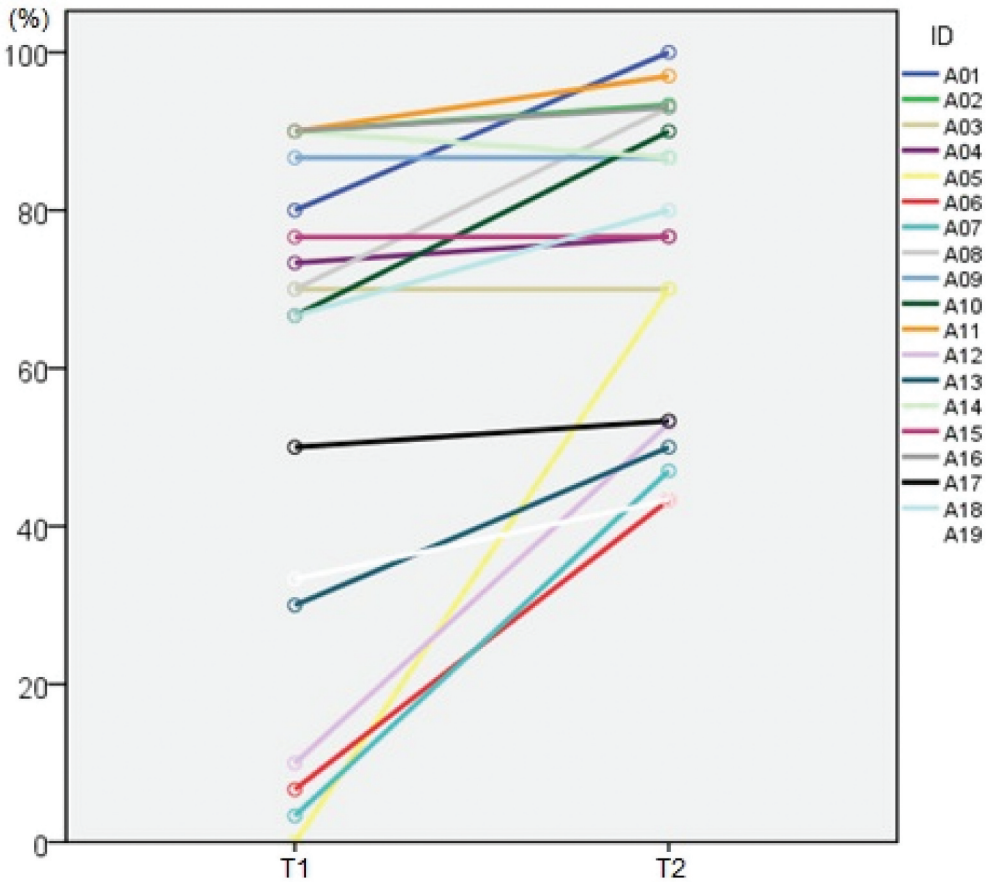
Figure 1. Individual changes as measured by Boston Naming Test Note. T1 = Acute stage; T2 = Chronic stage

Discussion

We investigated whether natural language recovery from acute to chronic stages is manifested differentially by tasks (word-level task vs discourse-level task), and the degree to which word retrieval ability at the word level is correlated with word retrieval ability at the discourse level at the two stages (acute & chronic). Our findings suggest PWA's performance was significantly different between BNT and HANA (BNT > HANA), and was significantly better at the chronic stage (T2) than at the acute stage (T1). Interestingly, according to the results of McNemar's Test, CoreLex shows that the percentage of PWA whose performance was normal significantly improved from T1 to T2. Moreover, correlations between confrontation naming tasks and CoreLex were found only for the acute stage.

Findings demonstrated a significant main effect of task, but the significance was not observed between confrontation naming tasks and a discourse task. The findings are different from previous studies that report statistical significance between the confrontation naming and discourse-level tasks (Mayer & Murray, 2003; Pashek & Tompkins, 2002).
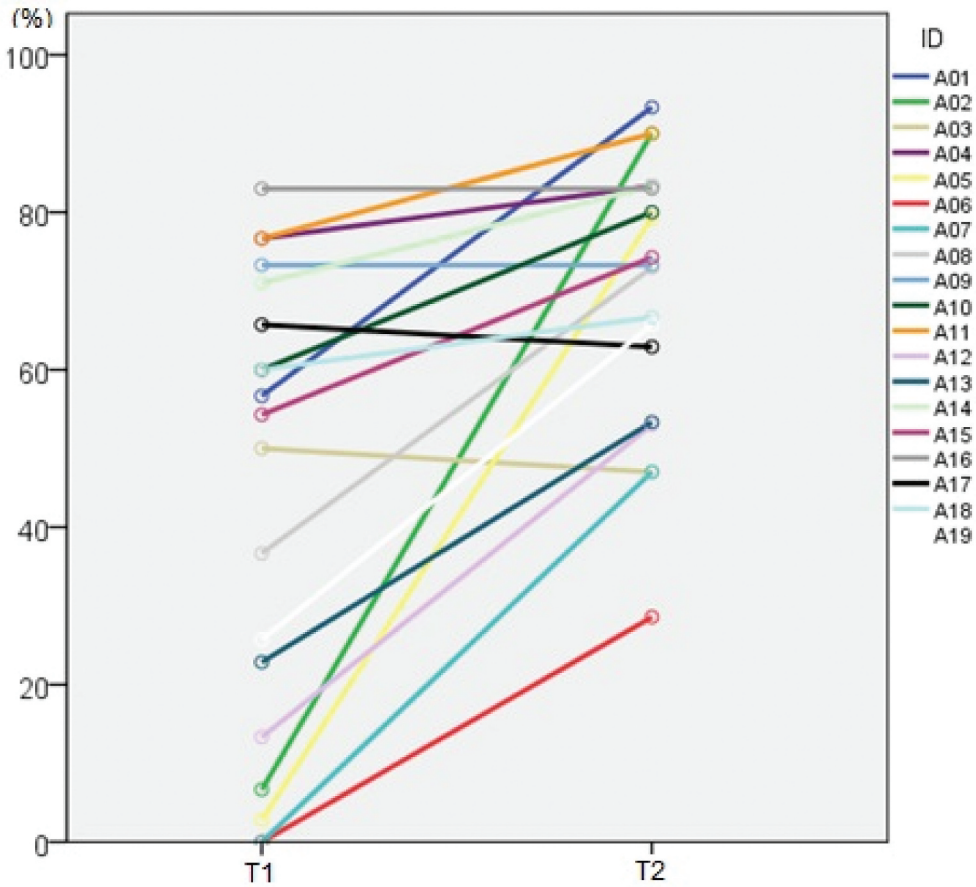
Figure 2. Individual changes as measured by Hopkins Action Naming Assessment Note. T1 = Acute stage; T2 = Chronic stage

The difference may be attributed to inherent properties of the discourse measure used in this current study as a tool to measure word retrieval ability at the discourse level. Mayer and Murray (2003) used percent word retrieval from the Test of Adolescent and Adult Word Finding (Gorman, 1990), percent substantive verbs, and percent corrected errors. Pashek and Thompkins (2002) used the number of target objects and actions in response to their stimuli. For CoreLex, scoring rules are stringent, only allowing target items for credit, and including few possible answers for patients to get credit, which is similar to confrontation naming tasks. Contrarily, it has been considered that scoring procedures of traditional discourse measures used in previous studies are more lenient compared to those of confrontation naming tasks (Law et al., 2018). Acceptable alternative responses to target objects or actions described in stimuli are not strictly pre-determined, and thus the range of possible correct responses can be broader. This also can be a possible reason that our patients demonstrate an opposite trend from previous studies, toward better performance on confrontation naming tasks than on the discourse task, though no significant difference was found. Although no difference was found between
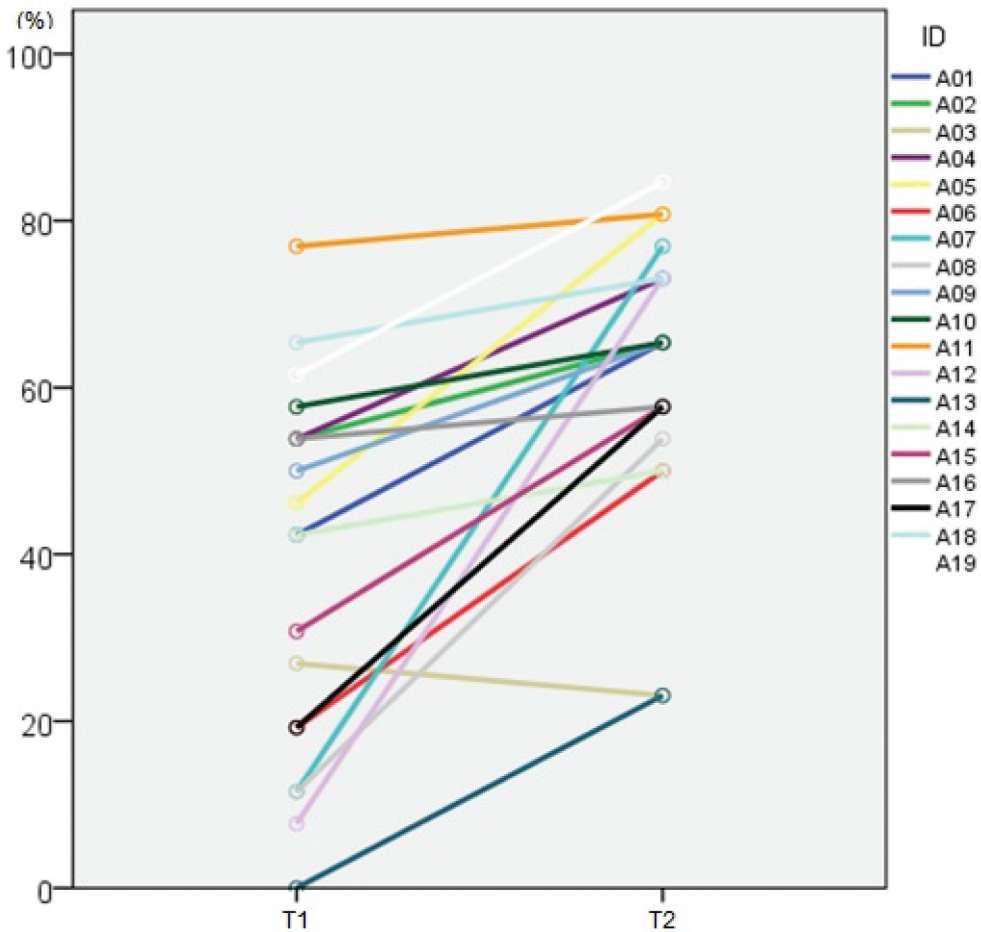
Figure 3. Individual changes as measured by CoreLex Note. T1 = Acute stage; T2 = Chronic stage

confrontation naming tasks and CoreLex using ANOVA, we additionally computed McNemar analyses to consider the clinical implications of the tasks. Our findings support the view that discourse-level assessment is an essential clinical procedure (e.g., Dietz & Boyle, 2018), especially to reveal improvements. This is because the absence of discourse-level assessments could have led to misinterpretation of the patients' language performance or recovery. Thus, interventions that have a significant effect on improving discourse might be inappropriately dismissed.

We found significant correlations between the two confrontation naming tasks at both time points and between naming accuracy and discourse performance acutely. Namely, high correlations were found between BNT and HANA in both the acute and chronic phases, and the correlations between picture naming for objects/actions and CoreLex was moderate at the acute time point. These findings are not surprising because several previous investigations have shown a broad range of correlations between discourse performance and confrontation naming performance (Boucher et al., 2022.; Fergadiotis &

Wright, 2016; Hillis et al., 2018; Mayer & Murray, 2003; Pashek & Tompkins, 2002; Richardson et al., 2018). The timing of assessments may account for the variability in time post-onset at which participants were tested. Our results indicate that confrontation naming tasks correlate with word retrieval ability at the discourse level at the acute stage, when much of the discourse might be limited to labelling items in the picture to be described. The correlation might be lower in the chronic phase because patients may produce more interpretive and integrative remarks as their linguistic abilities improve. Rather than label-ling objects and pictures, they may describe more accurate, complete sentences that assimilate the activities depicted in the picture (e.g., "because mom is not paying attention, her son may get hurt"). This study is the first report on the different patterns of correlational results at the acute and chronic stages. However, our interpretation is speculative.

Critically, these results should be interpreted within the context of the holistic approach to aphasia assessment. A lack of correlations at the chronic stage provides evidence that language recovery would not be manifested in all levels of language assessment in parallel. It is not possible to predict how language recovery seen in confrontation naming tasks translates to improvement in more naturalistic situations. As emphasized by previous studies (e.g., Fergadiotis & Wright, 2016), using confrontation naming tasks as a lens through which to make a clinical inference on discourse-level performance is likely to result in erroneous conclusions about patients' performance. Overall, our results imply that language recovery in aphasia should be measured at the discourse level, as reliance on confrontation naming tasks may not capture important gains in communication.

Findings regarding individual raw scores are in line with DeDe and Hoover's (2021) study reporting the potential feasibility of CoreLex. Participants (1 mild aphasia, 1 severe aphasia) in the study had conversation treatment, and their linguistic gains were quanti-fied using multiple discourse measures. They found that for the individual with mild aphasia, CoreLex showed an increase following treatment while other measures (e.g., percent correct information units) did not change. DeDe and Hoover attributed this finding to the potential impact of aphasia severity on treatment goals. Possibly, PWA may have achieved different goals through intervention by severity. It is likely that persons with mild aphasia perform at or near ceiling on naming tasks at the acute stage, and thus have restricted ranges of potential improvement in the confrontation naming task at the chronic stage. Individuals with mild aphasia may have targeted improvements in the context of naturalistic communication, which would not be targeted for individuals with severe profiles of aphasia. Tailoring treatment goals and protocols by severity of language ability is often reported in studies targeting the improvement of more functional linguistic skills (e.g., conversational skills) (Fox et al., 2009; Hoover et al., 2021; Kagan et al., 2001). Collectively, it is possible that utilization of discourse-level word retrieval tools like CoreLex might capture higher levels of language recovery, and are a more sensitive measure to detect communication gains that are not easily revealed in other measures, especially for mild aphasia.

Conclusions and limitations

Our findings have important implications for assessment tools used in clinical settings. By using CoreLex, linguistic changes can emerge in PWA, which often are not revealed in confrontation naming tasks. From a clinical perspective, if the intent is to longitudinally

track the recovery of aphasia, then the use of discourse outcome measures would be most sensitive and appropriate. Moreover, correlational results demonstrated language recovery is captured differentially by task at the chronic stage. This reflects that recovery patterns captured by discourse tasks is not in parallel with those captured by confrontation naming tasks. Given the increasing emphasis on ecologically valid assessment, CoreLex may be a valuable addition to clinical evaluation by offering an accurate snapshot of PWA's ability beyond the single-word level.

Our study has a few notable limitations. Since we have comparatively small number of PWA by subtypes, it was not possible to investigate differential magnitude of the relationship between word retrieval ability in context and in isolation by subtypes, shown in Richardson and colleagues' (2018) study. Although we mentioned that discourse measures may be more sensitive to capture changes in language function in persons with mild aphasia (e.g., anomia), our sample was not large enough to draw a firm conclusion about the differential sensitivity of the measure by subtypes. Further work should include larger sample sizes with a more balanced number of aphasia subtypes. Also, we unfortunately do not have information on whether our participants received any speech and language services, and if so, what intervention programs they were involved in. The addition of this information shed light on whether interventions contribute to changes at the chronic stage, and if the measure is suitable for measuring gradations of changes induced by intervention in clinical settings. Although the clinical implication of CoreLex holds great potential for evaluating changes in language in PWA over time, further studies are needed to establish clinical usability. As emphasized by Boyle (2014), assessment tools that involve higher linguistic and cognitive ability tend to entail session-to-session variabilities by their nature, which is often likely to bring out spurious test results. As a new language assessment, future studies should demonstrate stability of CoreLex before it is clinically applied to measure word retrieval ability in discourse.

ORCID

Hana Kim 🆔 http://orcid.org/0000-0002-4183-4867
Argye E. Hillis 🆔 http://orcid.org/0000-0002-5192-1171

Data availability statement

Data are available upon reasonable request from the authors.

References

Alary Gauvreau, C., Le Dorze, G., Croteau, C., & Hallé, M.-C. (2019). Understanding practices of speech-language pathologists in aphasia rehabilitation: a grounded theory study. *Aphasiology*, *33*(7), 846–864.

Boucher, J., Marcotte, K., Brisebois, A., Courson, M., Houzé, B., Desautels, A., . . . & Brambati, S. M. (2022). Word-finding in confrontation naming and picture descriptions produced by individuals with early post-stroke aphasia. *The Clinical Neuropsychologist*, *36*(6), 1422–1437.

Boyle, M. (2014). Test–retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research*, *57*(3), 966–978.

Breining, B L, Tippett, D. C., Davis, C., Posner, J., Sebastian, R., Oishie, K., & Hillis, A. E. (2015, May). *Assessing dissociations of object and action naming in acute stroke* [poster Presentation]. Clinical Aphasiology Conference, Monterey, CA.

Breining, Bonnie L, Faria, A. V, Caffo, B., Meier, E. L., Sheppard, S. M., Sebastian, R., Tippett, D. C., & Hillis, A. E. (2021). Neural regions underlying object and action naming: complementary evidence from acute stroke and primary progressive aphasia. *Aphasiology*, 1–29.

Dalton, S. G. H., Kim, H., Richardson, J. D., & Wright, H. H. (2020). A Compendium of Core Lexicon Checklists. *Seminars in Speech and Language*, *41*(1), 41–60. https://doi.org/10.1055/s-0039-3400972

Dalton, S. G., Hubbard, H. I., Al Harbi, M., Berube, S., Apple, K., & Lynch, V. (2021). *Comparison of Main Concept and Core Lexicon Productions Between the Modern and Original Cookie Theft Stimuli in Healthy Control Participants*. Proceedings of the Academy of Aphasia, 6358.

Dalton, S. G., & Richardson, J. D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. *American Journal of Speech-Language Pathology*, *39*(11), 1125–1137. https://doi.org/10.1044/2015_AJSLP-14-0161

DeDe, G., & Hoover, E. (2021). Measuring change at the discourse-level following conversation treatment: Examples from mild and severe aphasia. *Topics in Language Disorders*, *41*(1), 5–26.

Dietz, A., & Boyle, M. (2018). Discourse measurement in aphasia research: have we reached the tipping point? *Aphasiology*, *32*(4), 459–464.

Fergadiotis, G., Kapantzoglou, M., Kintz, S., & Wright, H. H. (2019). Modeling confrontation naming and discourse informativeness using structural equation modeling. *Aphasiology*, *33*(5), 544–560. https://doi.org/10.1080/02687038.2018.1482404

Fergadiotis, G., Kellough, S., & Hula, W. D. (2015). Item response theory modeling of the Philadelphia Naming Test. *Journal of Speech, Language, and Hearing Research*, *58*(3), 865–877.

Fergadiotis, G., & Wright, H. H. (2016). Modelling confrontation naming and discourse performance in aphasia. *Aphasiology*, *30*(4), 364–480. https://doi.org/10.1080/02687038.2015.1067288

Fox, S., Armstrong, E., & Boles, L. (2009). Conversational treatment in mild aphasia: A case study. *Aphasiology*, *23*(7–8), 951–964.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *BDAE-3: Boston Diagnostic Aphasia Examination–Third Edition*. Philadelphia, PA: Lippincott Williams & Wilkins.

Gorman, D. J. (1990). T*he Test of Adolescent and Adult Word-Finding*. Austsin, TX: Pro-Ed.

Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology*, *22*(2), 184–203.

Hillis, A. E. (2007). Aphasia: progress in the last quarter of a century. *Neurology*, *69*(2), 200–213.

Hillis, A. E., Beh, Y. Y., Sebastian, R., Breining, B., Tippett, D. C., Wright, A., Saxena, S., Rorden, C., Bonilha, L., Basilakos, A., Yourganov, G., & Fridriksson, J. (2018). Predicting recovery in acute poststroke aphasia. *Annals of Neurology*, *83*(3), 612–622. https://doi.org/10.1002/ANA.25184

Hoover, E., DeDe, G., & Maas, E. (2021). A Randomized controlled trial of the effects of group conversation treatment on monologic discourse in Aphasia. *Journal of Speech, Language, and Hearing Research*, *64*(12), 4861–4875.

Kagan, A., Black, S. E., Duchan, J. F., Simmons-Mackie, N., & Square, P. (2001). Training volunteers as conversation partners using" Supported Conversation for Adults With Aphasia"(SCA): A controlled trial. *Journal of Speech, Language, and Hearing Research*, *44*(3), 624–638.

Kaplan, E. F., Goodglass, H., & Weintraub, S. (2001). *The Boston Naming Test (2nd ed.)*. Philadelphia, PA: Lippincott Williams & Wilkins

Keator, L. M., Faria, A. V, Kim, K. T., Saxena, S., Wright, A. E., Sheppard, S. M., Breining, B. L., Goldberg, E., Tippett, D. C., & Meier, E. (2020). An efficient bedside measure yields prognostic implications for language recovery in acute stroke patients. *Cognitive and Behavioral Neurology*, *33*(3), 192–200.

Kertesz, A. (2006). *Western Aphasia Battery–Revised (WAB-R)*. Austin, TX: Pro-Ed.

Kim, H., Kintz, S., & Wright, H. H. (2021). Development of a measure of function word use in narrative discourse: Core lexicon analysis in aphasia. *International Journal of Language & Communication Disorders*, *56*(1), 6–19.

Kim, H., Kintz, S., Zelnosky, K., & Wright, H. H. (2019). Measuring word retrieval in narrative discourse: core lexicon in aphasia. *International Journal of Language & Communication Disorders*, *54*(1), 62–78.

Kim, H., Schoemann, A. M., & Wright, H. H. (2022). Quality of measurement in core lexicon measures. *Journal of Speech, Language, and Hearing Research*, *65*(8), 2919–2930.

Kim, H., & Wright, H. H. (2020a). Concurrent validity and reliability of the core lexicon measure as a measure of word retrieval ability in aphasia narratives. *American Journal of Speech-Language Pathology*, *29*(1), 101–110.

Kim, H., & Wright, H. H. (2020b). A tutorial on core lexicon: development, use, and application. *Seminars in Speech and Language*, *41*(1), 20–31.

Kiran, S., Cherney, L. R., Kagan, A., Haley, K. L., Antonucci, S. M., Schwartz, M., Holland, A. L., & Simmons-Mackie, N. (2018). Aphasia assessments: A survey of clinical and research settings. *Aphasiology*, *32*(S1), 47–49.

Law, S. P., Kong, A. P. H., & Lai, C. (2018). An analysis of topics and vocabulary in Chinese oral narratives by normal speakers and speakers with fluent aphasia. *Clinical Linguistics & Phonetics*, *32*(1), 88–99.

Long, C., Sebastian, R., Faria, A. V, & Hillis, A. E. (2018). Longitudinal imaging of reading and naming recovery after stroke. *Aphasiology*, *32*(7), 839–854.

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology*, *24*(6–8), 856–868. https://doi.org/10.1080/02687030903452632

Mason, C., & Nickels, L. (2022). Are single-word picture naming assessments a valid measure of word retrieval in connected speech?. *International Journal of Speech-Language Pathology*, *24*(1), 97–109.

Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology*, *17*(5), 481–497.

Pashek, G. V., & Tompkins, C. A. (2002). Context and word class influences on lexical retrieval in aphasia. *Aphasiology*, *16*(3), 261–286. https://doi.org/10.1080/02687040143000573

Richardson, J. D., Grace, S., Dalton, H., Fromm, D., Forbes, M., Holland, A., & Macwhinney, B. (2018). The Relationship Between Confrontation Naming and Story Gist Production in Aphasia. *American journal of speech-language pathology*, *27*, 406–422.

Simmons-Mackie, N., & Damico, J. S. (2010). Exploring clinical interaction in speech-language therapy: Narrative, discourse and relationships. *Therapeutic processes for communication disorders*. pp. 51–68.