ASHA American Speech-Language-Hearing Association

JSLHR

Research Note

Introducing the NEURAL Research Lab Data Set for Studies of Discourse and Gesture in Aphasia and Cognitively Healthy Aging Adults

Brielle C. Stark,[a] Kent Meinert,[a] Katelyn Urena,[a] Grace Oeding,[a] Davida Fromm,[b] and Brian MacWhinney[b]

[a] Indiana University Bloomington [b] Carnegie Mellon University, Pittsburgh, PA

ARTICLE INFO

Article History:
Received October 17, 2024
Revision received March 13, 2025
Accepted July 9, 2025

Editor-in-Chief: Julie A. Washington
Editor: William S. Evans

https://doi.org/10.1044/2025_JSLHR-24-00732

ABSTRACT

Purpose: The NEURAL Research Lab data set was collected with the intention of improving the study of short-term variation in spoken language and co-speech gesture in persons with aphasia as well as cognitively healthy adults. The purpose of this short note is to introduce readers to the available data and present two use cases as examples of leveraging the data set.
Method: The final data set will include a minimum of 75 adults with aphasia and 75 cognitively healthy adults, tested across two time points with a short time (about a week) between sessions. The focus was on capturing spoken discourse and manual gesture.
Results: For the purpose of demonstrating the data's potential, we examine two case studies: (a) evaluating test-retest reliability of fluency metrics derived from narrative spoken discourse and (b) characterizing communicative gesture use during spoken discourse.
Conclusions: In this short research note, we methodically introduce a novel open data set, NEURAL-2, available on AphasiaBank, for the examination of naturalistic gesture and speech in persons with and without aphasia. The first example use case analysis provided evidence that persons with mildest aphasia are less fluent than cognitively healthy peers, despite being younger, and that fluency metrics appear to have good test–retest reliability. The second use case showed that gesture rates do not significantly differ between individuals with aphasia and cognitively healthy adults, do not relate to age or cognitive status, and are negatively related to aphasia severity. We intend to validate these analyses in the final large sample. Clinicians and researchers are encouraged to use this data set to improve the understanding of speech, language, and gesture in aphasia and aging.
Supplemental Material: https://doi.org/10.23641/asha.30203545

In individuals with aphasia, a language disorder often resulting from brain injury or stroke, discourse production can be severely impaired, leading to challenges in fluency, word retrieval, and sentence structure (Andreetta et al., 2012; Dalton & Richardson, 2019; Fromm et al., 2016; Nicholas & Brookshire, 1993; Stark et al., 2023). While discourse—language beyond a single phrase used for a specific purpose (Armstrong, 2000)—is central to communication, effective communication extends beyond spoken language, incorporating both verbal and nonverbal elements. Gestures, such as hand movements, also play a vital role in supporting and enhancing spoken language (Kendon, 1980; McNeill, 1992). Research has shown that many individuals with aphasia retain the ability to produce gestures, particularly those that are semantically related to speech, making gesture a valuable compensatory tool (Akhavan et al., 2018; Cocks et al., 2018; de Kleine et al., 2023; Dipper et al., 2015; Rose et al., 2017; Stark & Oeding, 2024; van Nispen et al., 2017).

Understanding how both spoken discourse and gesture production vary over time is crucial for advancing multimodal models of communication and for designing effective clinical assessments. While research suggests that discourse performance is influenced by task structure and elicitation

Correspondence to Brielle C. Stark: bcstark@iu.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

procedures in persons with aphasia (Dalton & Richardson, 2019; Deng et al., 2024; Dipper et al., 2018; Fergadiotis & Wright, 2011; Stark, 2019; Stark & Fukuyama, 2020) and cognitively healthy adults (CHAs; Fergadiotis et al., 2011; Stark & Fukuyama, 2020; Wright et al., 2013), less is known about the systematic reliability of discourse and gesture across repeated sampling (Boyle, 2014, 2015; Brookshire & Nicholas, 1994; Stark et al., 2023). Similarly, while aging is known to impact spoken discourse (Chapman et al., 2002; Glosser & Deser, 1992; Kim et al., 2024; C.-O. Martin et al., 2018; Wright et al., 2013), spontaneous gesture production and its stability over time remain underexplored in older adults (Arslan & Goksun, 2022; Ozer et al., 2017; Theocharopoulou et al., 2015).

This research note introduces NEURAL-2, a publicly available data set designed to support open science initiatives by providing systematically collected test–retest data on both spoken discourse and gesture in adults with aphasia and CHAs. The data set builds on prior work (NEURAL data set; Stark et al., 2023) and represents one of the few open-access resources of clinical and nonclinical populations evaluating language and gesture in a controlled, short-term test–retest design. By making these data freely available, we aim to provide researchers and clinicians with resources to investigate variability in discourse and gesture production over time, inform methodological best practices, improve clinical education with multimodal examples and speech across a variety of tasks, and foster reproducible research in communication sciences.

Unlike traditional research reports that emphasize hypothesis-driven analyses, this research note is intended to facilitate broad reuse of the data set. To illustrate potential applications, we present two use cases/analyses: (a) evaluating test–retest reliability of fluency features during narrative spoken discourse and (b) characterizing communicative gesture production across aging and aphasia. These analyses serve as demonstrations of data set utility rather than definitive findings. We encourage researchers and clinicians to engage with NEURAL-2 (https://aphasia.talkbank.org/access/English/Protocol/NEURAL-2.html) to address open questions in discourse and gesture research.

Method and Design

Participant Inclusion and Exclusion Parameters

The current study was approved by the Indiana University Institutional Review Board (14633), and informed consent was obtained using the REDCap e-Consent framework (Harris et al., 2009, 2019). There were two groups: an Aphasia group and a CHA group. All participants self-reported adequate hearing and vision, either unaided or aided (e.g., glasses, hearing aids). They were fluent English speakers with no self-reported neurodegenerative conditions such as Alzheimer's or Parkinson's disease.

The Aphasia group included individuals aged 18–85 years who had acquired a brain injury (e.g., stroke, traumatic brain injury) at least 6 months prior and who had received a diagnosis of aphasia. Notably, the study included individuals with latent aphasia, defined as having a known brain injury and self-reported language impairments, regardless of standardized aphasia cutoffs (DeDe & Salis, 2020; Silkes et al., 2021; Stark et al., 2025).

The CHA group included individuals aged 30–85 years, an age range chosen to mirror the anticipated Aphasia group. In the first wave of data collection for this group, we focused on recruiting individuals over 55 years of age, given the limited availability of spoken discourse and gesture data sets for this demographic. Participants had no self-reported history of neurological conditions, including stroke or head injury. Those scoring below 23 (out of 30) on the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) were excluded.

Design

Data were collected using a short-term test–retest design, with 5–10 days between administrations of test (Day 1) and retest (Day 2; same design as in Stark et al., 2023). For the Aphasia group, because it took longer to complete all assessments (see Table 1 for a complete list and description of assessments), both test (Day 1) and retest (Day 2) were split into 2-hr morning (a.m.) and afternoon (p.m.) sessions. For the CHA group, test (Day 1) and retest (Day 2) were completed in two half days, usually 60–90 min in length. Following virtual testing considerations featured in previous work (Doub et al., 2021; Stark et al., 2023), all procedures occurred virtually via Zoom Health. For speaking tasks, the participant was asked to move away from the computer so that the entire gesture space was in view, such that any gesture of the hands was easily observed (Stark et al., 2021). This also prevented the participant from resting their hands on a table or desk, which may have impacted their tendency to gesture.

Data Collected

This study is part of a collaborative project between Indiana University and Carnegie Mellon University. The data collection procedures occur at Indiana University, as do the preliminary data preparation and validation. A final data check and then curation occur at Carnegie Mellon University.

Table 1 lists the neuropsychological and discourse tasks collected during the study. Briefly, participants

Table 1. All tasks collected during the study, including a short description, and the participant groups.

Task name	Task description	Test/Day 1		Retest/Day 2		Part of original AphasiaBank protocol?
		Aphasia	CHA	Aphasia	CHA	
Informed consent and intake	Participants provided informed consent and completed an intake form about demographic information.	Yes (a.m.)	Yes	Not collected	Not collected	Yes
Western Aphasia Battery–Revised, Part I	The Western Aphasia Battery–Revised, Part I, was administered to assess aphasia severity, scored out of a possible 100.	Yes (a.m.)	Not collected	Not collected	Not collected	Yes
Montreal Cognitive Assessment[a]	Validated screening tool for early cognitive impairment, scored out of a possible 30	Not collected	Yes	Not collected	Not collected	Yes
Patient Health Questionnaire–9	Self-report 9-item screener meant to objectify and assess degree of depression severity, ranging from 0 (*no indication of depression*) to 27 (*severe depression*)	Yes (a.m.)	Yes	Not collected	Not collected	No (self-report only)
Test of upper limb apraxia	12-item screener meant to identify upper limb apraxia through the use of imitation and pantomime	Yes (a.m.)	Not collected	Not collected	Not collected	No
Discourse battery	Several speaking samples were acquired: 1. Dialogue and free speech a. 2-min dialogue on particular topic(s)^ b. Free speech on a prompt (“How do you think your speech is these days?) 2. Narratives a. Stroke Story or Illness Story b. Important Event c. Neutral Cue^ 3. Picture descriptions/exposition a. Cat Rescue (single picture) b. Broken Window (four-pane sequence) c. Refused Umbrella (six-pane sequence) 4. Fictional narrative a. Cinderella 5. Procedural narratives a. Sandwich b. Laundry^ c. Flower^ (only administered at retest [Day 2]) d. Getting dressed^ (only administered at test [Day 1])	Yes (p.m.)	Yes	Yes (p.m.)	Yes	Yes (partially; some new tasks added, marked with ^)

(*table continues*)

Table 1. *(Continued).*

Task name	Task description	Test/Day 1		Retest/Day 2		Part of original AphasiaBank protocol?
		Aphasia	CHA	Aphasia	CHA	
Language Experience and Proficiency Questionnaire	Validated questionnaire tool for collecting self-reported proficiency and experience data regarding known languages	Not collected	Yes	Not collected	Not collected	No
Temple Assessment of Language and Short-term Memory in Aphasia	Validated test battery to assess language and verbal short-term memory abilities in aphasia. The online, freely available web version was used.	Not collected	Not collected	Yes (a.m.)	Not collected	No
Philadelphia Naming Test (Short Form)	Short form (A) was used, containing 30 black and white pictures of objects to be named.	Not collected	Not collected	Yes (AM)	Not collected	No (Boston Naming Test used)
Verb Naming Test	22 pictures showing an action to be named, from the Northwestern Assessment of Verbs and Sentences–Revised	Not collected	Not collected	Yes (a.m.)	Not collected	Yes
Concrete and abstract picture description task	50 different images that participants are asked to describe, to examine how individuals use speech and gesture, to describe abstract vs. concrete Timages	Not collected	Not collected	Yes	Yes	No
Address description task	A “procedural” task to examine how individuals with aphasia and age-matched controls use speech and gesture to describe how to get from one familiar point to another	Not collected	Not collected	Yes	Yes	No
Wechsler Adult Intelligence Scale–Fourth Edition	Validated adult measure of cognition. Only the subtests that contributed to the Working Memory Index were acquired.	Not collected	Not collected	Not collected	Yes	No

Note. It is indicated if task was collected during Session 1 (a.m.) or Session 2 (p.m.) for Aphasia group. Whether the tasks are part of the original AphasiaBank protocol is noted. CHA = cognitively healthy adult.

[a]Exclusionary metric (score < 23 excluded).

provided demographic data (e.g., age, sex, education, handedness, language status) at intake, followed by the Patient Health Questionnaire–9 (PHQ-9) to assess self-reported depression (Kroenke et al., 2001). Cognitive and language assessments varied by group, but with the goal of collecting comparable metrics between groups. For the Aphasia group, the Western Aphasia Battery–Revised (WAB-R) Part I (Kertesz, 2007) was administered to assess type and severity of aphasia. WAB-R scoring was verified by author D.F. or B.C.S. by evaluating the video of the session. Additional language measures included the short-form Philadelphia Naming Test (PNT; Walker & Schwartz, 2012) and the Verb Naming Test (Cho-Reyes & Thompson, 2012). The short PNT was selected over the Boston Naming Test (which is the typical test used in AphasiaBank protocols) because of the burgeoning research using the PNT to refine psycholinguistic models (e.g., Dell et al., 2007, 2013; Schwartz et al., 2004) and understand error making (e.g., Casilio et al., 2023; Schwartz et al., 2004). Verbal short-term memory (digit span and letter span) was evaluated using the Temple Assessment of Language and Short-Term Memory in Aphasia (TALSA; N. Martin et al., 2018), although technical issues resulted in missing data for approximately 20 participants. The TALSA was collected because of its specialization for measuring working memory in aphasia. Motor and discourse-based assessments were also conducted for the Aphasia group. The test of upper limb apraxia (Vanbellingen et al., 2010) screened for motor impairments potentially affecting gesture production.

For the CHA group, the MoCA (Nasreddine et al., 2005) was administered in addition to the PHQ-9 and the demographic questionnaire. Additional assessments included subtests comprising the Working Memory Index from the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV; Wechsler, 2008; digit span, letter–number sequencing, and arithmetic) and the Language Experience and Proficiency Questionnaire (Marian et al., 2007) for bilingual participants. It was our choice to use the TALSA to measure working memory in lieu of the WAIS-IV in the Aphasia sample, given changes to WAIS-IV administration, subtest scoring, and score normalization for neurological groups (Loring & Bauer, 2010).

Both groups then completed a battery of spoken discourse tasks, adapted from the AphasiaBank protocol (MacWhinney et al., 2011). Some discourse tasks were part of the original AphasiaBank protocol, including free speech, two autobiographical narratives (Stroke [Aphasia] or Illness [CHA] story; Important Event), picture description ("Cat Rescue"), picture sequence expositions (Broken Window; Refused Umbrella), fictional story retell (Cinderella), and a procedural narrative (how to make a sandwich). A variety of new discourse tasks were added to NEURAL-2, specifically a 2-min dialogue between the experimenter (who was a novel conversation partner for the participant) and the participant on a topic (e.g., recent trip, favorite TV show), four narratives that were elicited by prompting the participant to expound on a neutral cue word (clock, bird, hotel, restaurant; from Kurczek et al., 2013), and three new procedural narratives (how to do laundry, test [Day 1] and retest [Day 2]; how to grow a flower outside, retest [Day 2] only; how to get dressed in the morning during summer, test only [Day 1]). The new narrative and procedural tasks were added because of the bulk of work suggesting that they are gesture-rich (Chawla & Krauss, 1994; Holler & Wilkin, 2009; Jacobs & Garnham, 2007; Pritchard et al., 2015; Stark & Cofoid, 2022). The protocol, instructions, and materials for eliciting all discourse samples are available on the NEURAL-2 AphasiaBank (https://talkbank.org/aphasia/access/English/Protocol/NEURAL-2.html).

Data Curation

Discourse Protocol

All sessions were video-recorded, and discourse segments were processed using batchalign2 (Liu et al., 2023), an automatic speech recognition system designed to interface with Computerized Language ANalysis (CLAN; MacWhinney, 2018). Batchalign2 aligns transcripts to video at the utterance and word levels, creating a transcript in CHAT format, which allows for efficient manual correction and a large variety of automatic linguistic analyses in CLAN (e.g., mean length of utterance, propositional idea density). Each transcript underwent three levels of review following automatic generation: an initial check by a research assistant, a secondary check by a different research assistant, and a final verification by the research team at Carnegie Mellon University. Once finalized, CLAN's MOR command was applied to annotate morphological and grammatical structures before uploading transcripts and videos to AphasiaBank.

Reliability. Trained research assistants manually reviewed each transcript and video, correcting errors in speaker ID, revisions, repetitions, utterance delineation, paraphasias, and transcription (e.g., incorrect spelling or capitalization). Due to the long duration of this project (3 years thus far for a total of 5 years), research assistants ("coders") rotate. Thus far, seven coders have been employed on the project. To establish a base quality of coding, all coders were trained on two gold-standard checked transcripts (one CHA, one Aphasia) to 80% agreement, which was often met following consensus training and discussion with authors B.C.S. and D.F.

A base transcript was generated using automatic speech recognition from batchalign2. Then, two raters independently opened and edited the base transcript for correctness. For all files, raters watch the entire video while also reading through the transcript, changing whatever is necessary. Raters often had to fix capitalization, add codes (e.g., for paraphasias), add paraphasia targets, fix revisions and repetitions, and fix utterances so that they were a single communication unit (an independent clause with its modifiers). This was done for 14 CHA transcripts (25.92% of data collected to date) and 22 Aphasia transcripts (24.7% of data collected to date). We calculated interrater reliability using ICC(2, *k*), a two-way random-effects model for absolute agreement based on the average of two raters per transcript, with different rater pairs assigned across transcripts, reflecting the reliability of group mean ratings rather than individual consistency. Given the high quality of the base transcript, intraclass correlation coefficient (ICC) values were, overall, excellent for words (ICC = .999, 95% CI [0.998, 0.999]) and utterances (ICC = 1.00, 95% CI [0.999, 1.00]). The high interrater agreement held when evaluating each group separately (Aphasia: words, ICC = .999, 95% CI [0.997, 0.999]; utterances, ICC = .999, 95% CI [0.999, 1.00]; CHA: words, ICC = .998, 95% CI [0.995, 0.999]; utterances, ICC = 1.00, 95% CI [1.00, 1.00]).

For all other files in the sample, there was a primary coder and a checker. The primary coder did as described above, and then the checker used the primary coder's checked transcript to recheck.

Gesture Annotation Protocol

Author K.U. led a team of research assistants in manually coding co-speech gestures using EUDICO Linguistic Annotator (henceforth ELAN) software (ELAN, 2024; Lausberg & Sloetjes, 2016) and a converted ELAN2-CHAT file, which placed coded gestures into %tiers in the CHAT transcript. Our team is interested in capturing several components of communicative gestures—that is, those that are related to speech. Noncommunicative hand movements, such as body adjustments, are not coded. Our coding captures the following:

1. handedness (if gesture occurs with left, right, or both hands),
2. type (only representational gestures are coded at present, i.e., iconic, metaphoric, deictic, referential; McNeill, 1992),
3. handshape classifier (using American Sign Language classifiers, which are designated handshapes providing additional information about location, action, size, shape, and manner),
4. function (whether the gesture is supplementary to [adding, disambiguating, or replacing] speech or redundant [emphasizing, adding no additional information to] speech), and
5. for iconic gestures only, iconicity (handling, enacting, object, shape, or path; van Nispen et al., 2017).

The protocol for gesture coding is available on the AphasiaBank website, including more details on the aforementioned categories. See Figure 1 for a schematic of our coding.

Reliability. Interrater reliability for gesture coding is ongoing; a sample of n = 6 (11%) of the CHA data set has been coded for interrater reliability thus far. The three procedural discourse tasks collected at retest (Day 2) were coded (Sandwich, Laundry, Flower). Using ICC(2, 1) with absolute agreement, the total number of communicative gestures showed excellent agreement (ICC = .986, 95% CI

Figure 1. Example of gesture coding scheme.

Example 1:
- Gesture: Hand raises up over head
- Speech: "He's producing work at a high level"
- Coding Scheme: RED:MET (Redundant + Metaphoric)

Example 2:
- Gesture: Hand raises up over head
- Speech: "The first floor is higher than the second floor"
- Coding Scheme: RED:IC:PA (Redundant + Iconic + Path).

Example 3:
- Gesture: Pretending to grab something
- Speech "I love to grab the sales when I buy flower seeds"
- Coding Scheme: RED:MET (Redundant + Metaphoric)

Example 4:
- Gesture: Pretending to grab something
- Speech: "I grab the seeds and place them into the dirt"
- Coding Scheme: RED:IC:EN (Redundant + Iconic + Enacting)

[0.905, 0.998]). Additionally, we evaluated the coding of function (supplementary vs. redundant), also finding good-to-excellent agreement (ICC = .943, 95% CI [0.556, 0.992]). In general support of the strength of our lab's coding, two of the research assistants working on the NEURAL-2 data set have also coded gestures in a clinical data set, showing on average good-to-excellent interrater reliability (Urena & Stark, 2024).

Results

Participant Overview

In total, 89 participants with aphasia were tested, and data collection is now complete for this group. The demographics for this group are as follows: age, $M = 56$ years ($SD = 12$, range: 20–85); education, $M = 17$ years ($SD = 3$, range: 12–25); race (White, $n = 65$; Asian, $n = 10$; Black or African American, $n = 8$; Indigenous American/Alaskan Native, $n = 3$; more than one race, $n = 2$; unknown/not reported, $n = 1$); ethnicity (not Hispanic or Latino, $n = 84$; Hispanic or Latino, $n = 3$; unknown/not reported, $n = 2$); sex (47 female and 42 male); and language status (72 monolingual and 17 bi/multilingual). On average, the sample trends toward mild aphasia (WAB-R Aphasia Quotient [AQ] on a 100-point scale, $M = 84.13$, $SD = 19.62$, range: 0.2–100). See Supplemental Material S1 for data from all clinical participants.

At this point (March 2025), 54 CHA participants have been tested: age, $M = 63$ years ($SD = 11$, range: 30–80); education, $M = 17$ years ($SD = 2$, range: 12–25); race (White, $n = 47$; Black or African American, $n = 4$; more than one race, $n = 3$); ethnicity (not Hispanic or Latino, $n = 48$; Hispanic or Latino, $n = 5$; unknown/not reported, $n = 1$); sex (34 female and 20 male); and language status (41 monolingual, 13 bi/multilingual). Their MoCA scores were on a 30-point scale ($M = 27.27$, $SD = 2.14$, range: 23–30). The goal is a minimum of 75 CHA participants by the end of data collection (2 years from now). See Supplemental Material S2 for data from CHA participants collected to date.

Use Cases

The following analyses used Aphasia and CHA group data currently archived on AphasiaBank, as of March 3, 2025.

Example Analysis 1: Are fluency measures reliable? Fluency comprises many linguistic and speech components (Clough & Gordon, 2020; Fromm et al., 2024; Gordon & Clough, 2020). In Stark et al. (2023), "fluency" was measured using words per minute and was found to be highly reliable across discourse tasks (i.e., picture description, fictional narrative, procedural narrative) between test (Day 1) and retest (Day 2) for a pilot sample of persons with and without aphasia. Since then, the Carnegie Mellon team has been improving CLAN's FLUCALC command to evaluate fluency more comprehensively. With this in mind and with the knowledge that older age typically accompanies reductions in fluency (at least in reduction of words per minute and increase in word repetition; Beier et al., 2023), this use case example evaluates fluency reliability in more detail.

The following fluency metrics were extracted from the transcripts of the Cinderella story: words per minute (to corroborate our prior finding), percentage phrase repetitions, percentage word revisions, percentage phrase revisions, percentage phonological fragments, and percentage filled pauses (e.g., "uh," "um"; Fromm et al., 2024). CLAN version 05feb25 for Windows was used to run FLUCALC. Because the sample lengths were highly variable within and between groups, the fluency analysis was run on the first 100 words, and if a transcript had fewer than 100 words, the minimum allotted was 75 words. Furthermore, each participant had to have at least 75 words at both time points of Cinderella to be included. Because of this parameter, two participants were excluded from the Aphasia group (RC104: 72-year-old man, anomic aphasia, $AQ = 78.6$, 52 words at test [Day 1] and 64 at retest [Day 2]; RC254: 43-year-old man, Broca's aphasia, $AQ = 21.8$, 17 words at test [Day 1] and five at retest [Day 2]) and one from CHA group (RC209: 58-year-old man, MoCA = 26, 59 words at test [Day 1] and 64 at retest [Day 2]). See Table 2 for demographics of included participants; the final sample included $n = 35$ Aphasia and $n = 22$ CHA.

The participant groups were not statistically different in education (Mann–Whitney U test; $W = 1{,}612$, $p = .67$). The Aphasia group was significantly younger (Mann–Whitney U test; $W = 854$, $p < .0001$). Table 3 shows descriptive statistics and statistical significance (p values) for group and time point comparisons for fluency metrics of the Cinderella spoken discourse task. We first evaluated differences in fluency variables between the Aphasia and CHA groups, collapsing across time points. Nonparametric Mann–Whitney U tests were used, and p values were corrected for multiple comparisons using false discovery rate. The Aphasia group produced significantly fewer words per minute, a higher percentage of phonological fragments, a higher percentage of phrase repetitions, a higher percentage of word revisions, a higher percentage of phrase revisions, and a higher percentage of filled pauses (all corrected $p < .05$). We next evaluated test–retest reliability of metrics across all participants and within each subject group for the Cinderella discourse task. There were no significant differences between test

Table 2. Selected demographics and neuropsychological data for fluency discourse analysis.

Task name	Aphasia (*N* = 35) (*M*, *SD*, min–max)	CHA (*N* = 22) (*M*, *SD*, min–max)
Age (years)	57.9 (13.0, 27.0–85.0)	66.5 (8.41, 41.0–77.0)
Sex	34 female (48.6%) 36 male (51.4%)	28 female (63.6%) 16 male (36.4%)
Race/ethnicity	8 Asian (11.4%) 6 Black (8.6%) 54 White (77.1%) 2 Unreported (2.9%)	0 Asian (0%) 2 Black (4.5%) 42 White (95.5%)
Education (years)	17.5 (3.24, 12.0–24.0)	17.3 (3.30, 12.0–25.0)
Western Aphasia Battery–Revised, Part I (max score: 100)	88.6 (12.5, 42.9–99.2)	Not collected
Montreal Cognitive Assessment (max score: 30)	Not collected	27.7 (2.29, 23.0–30.0)
Types of aphasia	26 anomic (37.1%) 2 Broca's (2.9%) 4 conduction (5.7%) 32 latent/not aphasic by WAB (45.7%) 4 transcortical motor (5.7%) 2 Wernicke's (2.9%)	
Number of words on which the fluency metrics were calculated (from Cinderella sample), averaged across days	*M* = 98.99 *SD* = 4.10	*M* = 100 *SD* = 0

Note. CHA = cognitively healthy adult; WAB = Western Aphasia Battery.

(Day 1) or retest (Day 2) for any variable when evaluating across all participants (all $p > .96$; see Table 3). The findings were similar when looking within each group, in that there were no significant differences between test (Day 1) and retest (Day 2) for either the Aphasia (all $p > .96$) or the CHA (all $p > .96$) groups.

In summary, despite the Aphasia group presenting with predominantly mild or latent aphasia and being younger, they were less fluent than the CHA group during a story retell. This use case replicates our prior work, demonstrating that words per minute was impaired in aphasia and was reliable across a short test–retest timeframe for both groups, and extends these findings of impairment and reliability to five other metrics of fluency and to milder aphasia.

Example Analysis 2: Characterizing gesture across adult lifespan and in mild aphasia. Individuals with mild or latent aphasia retain more intact language abilities than those with moderate and severe aphasia (DeDe & Salis, 2020; Salis & DeDe, 2022; Silkes et al., 2021; Stark et al., 2025), although many with mildest aphasia perceive that their functional communication is impacted (Cavanaugh & Haley, 2020). While gesture frequency and rate are known to correlate with aphasia severity, where individuals with more severe aphasia and anomia tend to produce gestures at higher rates and which more often supplement speech (Kong et al., 2015; Sekine et al., 2013; Stark & Oeding, 2024), less is known about how mild or latent aphasia affects gesture production. These individuals likely retain the ability to produce semantically rich gestures, but the relationship between gesture and speech in this group has not been systematically examined or compared to cognitively healthy peers.

Similarly, aging affects spoken language, associating with phenomena such as tip-of-the-tongue states, but its impact on gesture use is less clear. Some studies suggest that older adults gesture less than younger individuals (Arslan & Goksun, 2022), whereas other work suggests that gesture may help resolve tip-of-the-tongue states but only in persons with smaller short-term memory capacity (Pyers et al., 2021). By comparing gestures in mild aphasia and cognitively healthy older adults and accounting for age, this use case provides insights into how subtle aphasia presentation and aging influence communicative gestures.

At present, our team has coded gestures in a subsample of participants, for one time point (n = 18 Aphasia; n = 26 CHA). Future work will evaluate gesture variability, but at present, we provide preliminary analyses to evaluate how gesture production during spontaneous speech is related to age *and* presence/absence of aphasia. This analysis focuses on communicative gestures produced during three procedural narratives, which are known to be rich in gesture and have been studied extensively in aphasia gesture research (Kistner et al., 2019; Pritchard et al., 2015; Sekine et al., 2013; Sekine & Rose, 2013; Stark & Cofoid, 2022; Stark & Oeding, 2024).

The group of persons with aphasia included in this analysis trended toward a mild and latent presentation (WAB-R AQ, M = 91.2, SD = 8.63, range: 71.3–99.4).

Table 3. Examining test–retest reliability for fluency variables extracted from the Cinderella spoken discourse samples.

	Test and retest collapsed			Groups collapsed			Test vs. retest, Aphasia			Test vs. retest, CHA		
Variable	**Aphasia, *M* (*SD*)**	**CHA, *M* (*SD*)**	***p* (Aphasia vs. CHA)**	**Test, *M* (*SD*)**	**Retest, *M* (*SD*)**	***p* (test vs. retest)**	**Aphasia test, *M* (*SD*)**	**Aphasia retest, *M* (*SD*)**	***p* (test vs. retest)**	**CHA test, *M* (*SD*)**	**CHA retest, *M* (*SD*)**	***p* (test vs. retest)**
Words per minute	67.07 (30.31)	122.18 (28.85)	< .0001*	86.53 (39.53)	90.16 (40.82)	.96	63.85 (28.53)	70.30 (32.07)	.96	122.61 (25.04)	121.75 (32.81)	.96
Phonological fragments (%)	2.35,(2.80)	0.44 (0.71)	< .0001*	1.64 (2.44)	1.59 (2.41)	.99	2.35 (2.84)	2.34 (2.79)	.99	0.50 (0.75)	0.38 (0.68)	.99
Phrase repetitions (%)	0.71 (1.00)	0.24 (0.47)	.003*	0.54 (0.84)	0.52 (0.90)	.96	0.73 (0.94)	0.70 (1.07)	.96	0.25 (0.54)	0.23 (0.41)	.96
Word revisions (%)	1.70 (1.46)	0.74 (1.05)	< .0001*	1.43 (1.45)	1.23 (1.34)	.96	1.75 (1.55)	1.65 (1.39)	.96	0.93 (1.13)	0.55 (0.94)	.96
Phrase revisions (%)	1.57 (1.37)	0.85 (0.82)	.001*	1.26 (1.08)	1.32 (1.38)	.96	1.59 (1.14)	1.56 (1.58)	.96	0.74 (0.73)	0.96 (0.92)	.96
Filled pauses (%)	13.33 (16.56)	3.64 (2.29)	< .0001*	9.83 (12.64)	9.36 (15.08)	.96	13.74 (14.84)	12.93 (18.33)	.96	3.60 (1.76)	3.68 (2.77)	.96

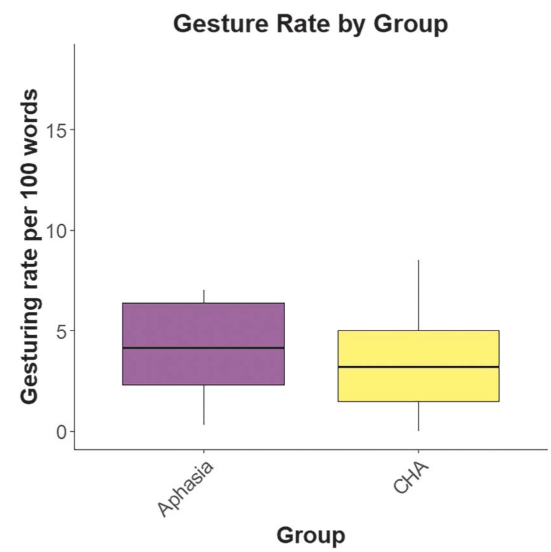
Note. *p* values have been corrected using false discovery rate. * indicates significant difference. CHA = cognitively healthy adult.

The CHA group's cognition scores (MoCA) were within typical limits based on race and education (M = 27.1 years, SD = 2.23, range: 23–30). Other important demographics of both groups include the following: sex (Aphasia: nine female, nine male; CHA: 14 female, 12 male), ethnicity (Aphasia: 5.6% Hispanic; CHA: 3.8% Hispanic), race (Aphasia: 72.2% White, 5.6% Black, 11.1% Asian, 5.6% Indigenous American/Alaska Native, 5.6% more than one race; CHA: 96.2% White, 3.8% Black, and 0% other categories), education (Aphasia: M = 18.1 years, SD = 3.88; CHA: M = 16.7 years, SD = 2.56), age (Aphasia: M = 57.1 years, SD = 11.8, range: 28–75; CHA: M = 65.4, SD = 11.6, range: 30–81), and languages spoken (Aphasia: 33.3% multilingual; CHA: 19.2% multilingual). The groups were not statistically different in education (Mann–Whitney U test, W = 277, p = .295), but the Aphasia group was significantly younger (Mann–Whitney U test, W = 131, p = .01).

Total communicative gestures produced across three procedural tasks (Sandwich, Laundry, Flower) from retest (Day 2) were analyzed for each participant. The average amount of verbal tokens produced across these tasks was M = 470 (SD = 355, range: 53–1,130) for the Aphasia group and M = 562 (SD = 274, range: 148–1,070) for the CHA group. The Aphasia group produced a total of 19.3 communicative gestures (SD = 21.5, range: 1–77), and the CHA group produced a total of 18.8 communicative gestures (SD = 16.2, range: 0–66). Given the heterogeneity in verbal tokens within and between groups, a standardized gesture rate variable was computed: number of gestures per 100 words (hereto forth, "gesture rate"). The Aphasia group's average gesture rate was 5.24 (SD = 4.74, range: 0.31–18.3), and the CHA group's rate was 3.39 (SD = 2.35, range: 0–8.49; see Figure 2).

There was not a significant difference in gesture rate between the Aphasia and CHA groups (Mann–Whitney U test, W = 288, p = .20). We evaluated a potential relationship between gesture rate and age across all participants, which was not significant ($r_s = -.29$, p = .058). We subsequently explored relationships between gesture rate with age *and* neuropsychological data (MoCA for CHA, WAB-R AQ for aphasia) for each group. Although the correlation between MoCA scores and gesture rate in the CHA group was not significant ($r_s = .36$, p = .07), age and gesture rate were significantly related ($r_s = -.47$, p = .015). We subsequently performed a partial correlation to evaluate whether age correlated with gesture rate after controlling for MoCA score because MoCA and age were also related in our sample, with older adults tending to score lower on the MoCA ($r_s = -.46$, p = .02). Age was no longer significantly related to gesture rate after controlling for MoCA score ($r_s = -.37$, p = .067). In the Aphasia group, there was a significant relationship between WAB-R AQ and gesture rate (lower AQ, more gesturing: $r_s = -.57$, p = .015), and age was not significantly related to gesture rate ($r_s = .08$, p = .75). In summary, this use case illustrates that age was not significantly related to gesture rate, even when considering neuropsychological information, but that aphasia severity, in our relatively mild sample, was.

Figure 2. A box and whisker plot that demonstrates the relationship between aphasia severity (as measured by Western Aphasia Battery–Revised Aphasia Quotient) and gesture rate per 100 words during three procedural narratives. This was a significant correlation.



Discussion and Conclusions

In this short note, we methodically introduce a novel open data set, NEURAL-2, available on AphasiaBank, for the examination of naturalistic gesture and speech in persons with and without aphasia. The first example use case analysis provided evidence that persons with mild and latent aphasia are less fluent than cognitively healthy peers, despite being younger, and that fluency metrics appear to have good test–retest reliability. The second use case showed that gesture rates do not significantly differ between individuals with aphasia and CHAs, do not relate to age or cognitive status, and are negatively related to aphasia severity. We intend to validate these analyses in the final large sample. We believe that our approach to data curation, combining automated tools (i.e., batchalign2) with detailed manual checks, establishes a robust framework for future analyses. The novel integration of detailed gesture annotations with established linguistic metrics using systems such as CLAN not only improves reliability but also sets the stage for more comprehensive multimodal analyses.

It is important to acknowledge the data set's limitations. Some data are incomplete due to technological interruptions (e.g., the TALSA website downtime). The relatively mild severity of aphasia represented in the sample may constrain the generalizability of findings to more severe aphasia but does widen the examination of subtle linguistic and gestural differences in mildest aphasia, which is a burgeoning area (DeDe & Salis, 2020; Fromm et al., 2017; Stark et al., 2025). At the time of writing this short note, the CHA population is not racially or ethnically representative of the United States. Author Urena received a Supplement to Promote Diversity to recruit a diverse sample for the to-be-collected CHA data. Another challenge is acquisition of spoken data via Zoom. Because the participant's internet quality cannot be controlled, this can result in speech that is subpar for fine-grained analyses (e.g., in Praat). In two cases, audio was sufficiently poor to warrant exclusion of the participant from the final data set. Finally, not all metrics were gathered at test (Day 1) and retest (Day 2). The AphasiaBank protocol metrics, which have time and again been shown to be a valuable tool for investigating various discourse metrics in clinical populations, were collected at test (Day 1) and retest (Day 2). Some new protocol additions, such as the procedural narratives, were collected only at test (Day 1) or retest (Day 2). The reason for this was primarily driven by participant experience, in that the testing sessions were already long (full days for the Aphasia group). While variability cannot be evaluated for all new tasks, collecting the new data provides opportunities for more comprehensive gesturing coding and evaluation of linguistic and discourse questions in new tasks (e.g., the neutral cue narratives). Our goal is to document these challenges so that researchers can accurately interpret the results of analyses conducted using our data set.

Clinicians and researchers are encouraged to use this data set to improve the understanding of speech, language, and gesture in aphasia and aging through research and clinical education. We also envision that this data set can be used for interdisciplinary purposes, for example, to improve artificial intelligence–driven models of automatic gesture identification and automatic speech recognition in clinical samples.

Data Availability Statement

As part of our commitment to transparency and open science, all data—including transcripts, videos, and gesture annotations—will be publicly available via AphasiaBank (https://aphasia.talkbank.org/), with additional data to be added as collection progresses. Some participants, due to their etiology, are archived in TBIBank (Traumatic Brain Injury cause, https://tbi.talkbank.org/) or RHDBank (Right Hemisphere Damage, https://rhd.talkbank.org/). AphasiaBank, TBIBank, and RHDBank membership is free for clinicians, researchers, and educators upon request.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant R01DC008524 awarded to Brian MacWhinney (Principal Investigator) and Brielle C. Stark (Co-investigator; "AphasiaBank: A Shared Database for the Study of Aphasic Communication"). Katelyn Urena was supported by a Supplement to Promote Diversity (2024–2026) on Grant R01DC008524. We would like to gratefully acknowledge each participant for lending us their valuable time. We would also like to acknowledge the study personnel who worked on transcription quality checking: Kennedy Skipper, Claire O'Dell, Isabel Visuri, Caroline Boman, Kali Manojlovich, Grace Oeding, Grace Rosswurm, and Madison Street.

References

Akhavan, N., Göksun, T., & Nozari, N. (2018). Integrity and function of gestures in aphasia. *Aphasiology, 32*(11), 1310–1335. https://doi.org/10.1080/02687038.2017.1396573

Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia, 50*(8), 1787–1793. https://doi.org/10.1016/j.neuropsychologia.2012.04.003

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology, 14*(9), 875–892. https://doi.org/10.1080/02687030050127685

Arslan, B., & Göksun, T. (2022). Aging, gesture production, and disfluency in speech: A comparison of younger and older adults. *Cognitive Science, 46*(2), Article e13098 https://doi.org/10.1111/cogs.13098

Beier, E. J., Chantavarin, S., & Ferreira, F. (2023). Do disfluencies increase with age? Evidence from a sequential corpus study of disfluencies. *Psychology and Aging, 38*(3), 203–218. https://doi.org/10.1037/pag0000741

Boyle, M. (2014). Test–retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research, 57*(3), 966–978. https://doi.org/10.1044/2014_JSLHR-L-13-0171

Boyle, M. (2015). Stability of word-retrieval errors with the AphasiaBank stimuli. *American Journal of Speech-Language Pathology, 24*(4). https://doi.org/10.1044/2015_AJSLP-14-0152

Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample size and test–retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research, 37*(2), 399–407. https://doi.org/10.1044/jshr.3702.399

Casilio, M., Fergadiotis, G., Salem, A. C., Gale, R. C., McKinney-Bock, K., & Bedrick, S. (2023). ParAlg: A paraphasia algorithm for multinomial classification of picture naming errors. *Journal of Speech, Language, and Hearing Research, 66*(3), 966–986. https://doi.org/10.1044/2022_JSLHR-22-00255

Cavanaugh, R., & Haley, K. L. (2020). Subjective communication difficulties in very mild aphasia. *American Journal of Speech-*

Language Pathology, 29(1S), 437–448. https://doi.org/10.1044/2019_AJSLP-CAC48-18-0222

Chapman, S. B., Zientz, J., Weiner, M., Rosenberg, R., Frawley, W., & Burns, M. H. (2002). Discourse changes in early Alzheimer disease, mild cognitive impairment, and normal aging. *Alzheimer Disease and Associated Disorders, 16*(3), 177–186. https://doi.org/10.1097/00002093-200207000-00008

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30*(6), 580–601. https://doi.org/10.1006/jesp.1994.1027

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250–1277. https://doi.org/10.1080/02687038.2012.693584

Clough, S., & Gordon, J. K. (2020). Fluent or nonfluent? Part A. Underlying contributors to categorical classifications of fluency in aphasia. *Aphasiology, 34*(5), 515–539. https://doi.org/10.1080/02687038.2020.1727709

Cocks, N., Byrne, S., Pritchard, M., Morgan, G., & Dipper, L. (2018). Integration of speech and gesture in aphasia. *International Journal of Language & Communication Disorders, 53*(3), 584–591. https://doi.org/10.1111/1460-6984.12372

Dalton, S. G., & Richardson, J. (2019). A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. *American Journal of Speech-Language Pathology, 28*(1S), 293–320. https://doi.org/10.1044/2018_AJSLP-17-0166

DeDe, G., & Salis, C. (2020). Temporal and episodic analyses of the story of Cinderella in latent aphasia. *American Journal of Speech-Language Pathology, 29*(1S), 449–462. https://doi.org/10.1044/2019_ajslp-cac48-18-0210

de Kleine, N., Rose, M. L., Weinborn, M., Knox, R., & Fay, N. (2023). Does gesture improve the communication success of people with aphasia?: A systematic review. *Aphasiology, 38*(3), 462–486. https://doi.org/10.1080/02687038.2023.2207781

Dell, G. S., Martin, N., & Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: Predicting word repetition from picture naming. *Journal of Memory and Language, 56*(4), 490–520. https://doi.org/10.1016/j.jml.2006.05.007

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Branch Coslett, H. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition, 128*(3), 380–396. https://doi.org/10.1016/j.cognition.2013.05.007

Deng, B.-M., Liang, L.-S., Zhao, J.-X., Zheng, H.-Q., & Hu, X.-Q. (2024). Correct information unit analysis in different discourse tasks among persons with anomic aphasia based on Mandarin AphasiaBank. *American Journal of Speech-Language Pathology, 33*(2), 800–813. https://doi.org/10.1044/2023_AJSLP-23-00217

Dipper, L., Pritchard, M., Morgan, G., & Cocks, N. (2015). The language–gesture connection: Evidence from aphasia. *Clinical Linguistics and Phonetics, 29*(8–10), 748–763. https://doi.org/10.3109/02699206.2015.1036462

Dipper, L., Pritchard, M., Walkden, E., & Cruice, M. (2018). How do speakers with and without aphasia use syntax and semantics across two discourse genres? *Aphasiology, 32*(6), 720–738. https://doi.org/10.1080/02687038.2018.1447642

Doub, A., Hittson, A., & Stark, B. C. (2021). Conducting a virtual study with special considerations for working with persons with aphasia. *Journal of Speech, Language, and Hearing Research, 64*(6), 2038–2046. https://doi.org/10.1044/2021_JSLHR-20-00392

ELAN. (2024). *ELAN* (Version 6.8) [Computer software]. https://archive.mpi.nl/tla/elan

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology, 25*(11), 1414–1430. https://doi.org/10.1080/02687038.2011.603898

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology, 25*(10), 1261–1278. https://doi.org/10.1080/02687038.2011.606974

Fromm, D., Chern, S., Geng, Z., Kim, M., Greenhouse, J., & MacWhinney, B. (2024). Automated analysis of fluency behaviors in aphasia. *Journal of Speech, Language, and Hearing Research, 67*(7), 2333–2342. https://doi.org/10.1044/2024_JSLHR-23-00659

Fromm, D., Forbes, M., Holland, A., & Dalton, G. (2017). Discourse characteristics in aphasia beyond the Western Aphasia Battery cutoff. *American Journal of Speech-Language Pathology, 26*(3), 762–768. https://doi.org/10.1044/2016_AJSLP-16-0071

Fromm, D., Greenhouse, J., Hou, K., Russel, G., Cai, X., Forbes, M., Holland, A., & MacWhinney, B. (2016). Automated proposition density analysis for discourse in aphasia. *Journal of Speech, Language, and Hearing Research, 59*(5), 1123–1132. https://doi.org/10.1044/2016_JSLHR-L-15-0401

Glosser, G., & Deser, T. (1992). A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. *Journal of Gerontology, 47*(4), P266–P272. https://doi.org/10.1093/GERONJ/47.4.P266

Gordon, J. K., & Clough, S. (2020). How fluent? Part B. Underlying contributors to continuous measures of fluency in aphasia. *Aphasiology, 34*(5), 643–663. https://doi.org/10.1080/02687038.2020.1712586

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics, 95,* Article 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research Electronic Data Capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24*(2), 267–289. https://doi.org/10.1080/01690960802095545

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*(2), 291–303. https://doi.org/10.1016/j.jml.2006.07.011

Kendon, A. (1980). Gesticulation and Speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–228). De Gruyter Mouton. https://doi.org/10.1515/9783110813098.207

Kertesz, A. (2007). *Western Aphasia Battery–Revised*. The Psychological Corporation.

Kim, H., Kintz, S., & Wright, H. (2024). Preliminary findings of expository discourse in older adults. *Journal of Language and Aging Research, 2*(2). https://doi.org/10.15460/jlar.2024.2.2.1325

Kistner, J., Marshall, J., & Dipper, L. T. (2019). The influence of conversation parameters on gesture production in aphasia. *Clinical Linguistics & Phonetics, 34*(8), 693–717. https://doi.org/10.1080/02699206.2019.1692075

Kong, A. P. H., Law, S. P., Wat, W. K. C., & Lai, C. (2015). Coverbal gestures among speakers with aphasia: Influence of aphasia severity, linguistic and semantic skills, and hemiplegia on gesture employment in oral discourse. *Journal of Communication Disorders, 56,* 88–102. https://doi.org/10.1016/j.jcomdis.2015.06.007

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9. *Journal of General Internal Medicine, 16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kurczek, J., Brown-Schmidt, S., & Duff, M. C. (2013). Hippocampal contributions to language: Evidence of referential processing deficits in amnesia. *Journal of Experimental Psychology: General, 142*(4), 1346–1354. https://doi.org/10.1037/a0034026

Lausberg, H., & Sloetjes, H. (2016). The revised NEUROGES–ELAN system: An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behavior Research Methods, 48*(3), 973–993. https://doi.org/10.3758/s13428-015-0622-z

Liu, H., MacWhinney, B., Fromm, D., & Lanzi, A. (2023). Automation of language sample analysis. *Journal of Speech, Language, and Hearing Research, 66*(7), 2421–2433. https://doi.org/10.1044/2023_JSLHR-22-00642

Loring, D. W., & Bauer, R. M. (2010). Testing the limits: Cautions and concerns regarding the new Wechsler IQ and Memory scales. *Neurology, 74*(8), 685–690. https://doi.org/10.1212/WNL.0b013e3181d0cd12

MacWhinney, B. (2018). *Tools for analyzing talk. Part 2: The CLAN program*. Erlbaum.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Martin, C.-O., Pontbriand-Drolet, S., Daoust, V., Yamga, E., Amiri, M., Hübner, L. C., & Ska, B. (2018). Narrative discourse in young and older adults: Behavioral and NIRS analyses *Neuroscience, 10,* Article 69. https://doi.org/10.3389/fnagi.2018.00069

Martin, N., Minkina, I., Kohen, F. P., & Kalinyak-Fliszar, M. (2018). Assessment of linguistic and verbal short-term memory components of language abilities in aphasia. *Journal of Neurolinguistics, 48,* 199–225. https://doi.org/10.1016/j.jneuroling.2018.02.006

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research, 36*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

Ozer, D., Tansan, M., Özer, E., Chatterjee, A., & Goksun, T. (2017). The effects of gesture restriction on spatial language in young and elderly adults. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society: Computational Foundations of Cognition* (pp. 1471–1476).

Pritchard, M., Dipper, L., Morgan, G., & Cocks, N. (2015). Language and iconic gesture use in procedural discourse by speakers with aphasia. *Aphasiology, 29*(7), 826–844. https://doi.org/10.1080/02687038.2014.993912

Pyers, J. E., Magid, R., Gollan, T. H., & Emmorey, K. (2021). Gesture helps, only if you need it: Inhibiting gesture reduces tip-of-the-tongue resolution for those with weak short-term memory. *Cognitive Science, 45*(1), Article e12914. https://doi.org/10.1111/cogs.12914

Rose, M. L., Mok, Z., & Sekine, K. (2017). Communicative effectiveness of pantomime gesture in people with aphasia. *International Journal of Language & Communication Disorders, 52*(2), 227–237. https://doi.org/10.1111/1460-6984.12268

Salis, C., & DeDe, G. (2022). Sentence production in a discourse context in latent aphasia: A real-time study. *American Journal of Speech-Language Pathology, 31*(3), 1284–1296. https://doi.org/10.1044/2022_AJSLP-21-00232

Schwartz, M., Wilshire, C., Gagnon, D., & Polansky, M. (2004). Origins of nonword phonological errors in aphasic picture naming. *Cognitive Neuropsychology, 21*(2–4), 159–186. https://doi.org/10.1080/02643290342000519

Sekine, K., & Rose, M. L. (2013). The relationship of aphasia type and gesture production in people with aphasia. *American Journal of Speech-Language Pathology, 22*(4), 662–672. https://doi.org/10.1044/1058-0360(2013/12-0030)

Sekine, K., Rose, M. L., Foster, A. M., Attard, M. C., & Lanyon, L. E. (2013). Gesture production patterns in aphasic discourse: In-depth description and preliminary predictions. *Aphasiology, 27*(9), 1031–1049. https://doi.org/10.1080/02687038.2013.803017

Silkes, J. P., Zimmerman, R. M., Greenspan, W., Reinert, L., Kendall, D., & Martin, N. (2021). Identifying verbal short-term memory and working memory impairments in individuals with latent aphasia. *American Journal of Speech-Language Pathology, 30*(1S), 391–406. https://doi.org/10.1044/2020_AJSLP-19-00105

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American Journal of Speech-Language Pathology, 28*(3), 1067–1083. https://doi.org/10.1044/2019_AJSLP-18-0265

Stark, B. C., Alexander, J. M., Hittson, A., Doub, A., Igleheart, M., Streander, T., & Jewell, E. (2023). Test–retest reliability of microlinguistic information derived from spoken discourse in persons with chronic aphasia. *Journal of Speech, Language, and Hearing Research, 66*(7), 2316–2345. https://doi.org/10.1044/2023_JSLHR-22-00266

Stark, B. C., Clough, S., & Duff, M. (2021). Suggestions for improving the investigation of gesture in aphasia. *Journal of Speech, Language, and Hearing Research, 64*(10), 4004–4013. https://doi.org/10.1044/2021_JSLHR-21-00125

Stark, B. C., & Cofoid, C. (2022). Task-specific iconic gesturing during spoken discourse in aphasia. *American Journal of Speech-Language Pathology, 31*(1), 30–47. https://doi.org/10.1044/2021_AJSLP-20-00271

Stark, B. C., Dalton, S. G., & Lanzi, A. M. (2025). Access to context-specific lexical-semantic information during discourse tasks differentiates speakers with latent aphasia, mild cognitive impairment, and cognitively healthy adults. *Frontiers in Human Neuroscience, 18,* Article 1500735. https://doi.org/10.3389/fnhum.2024.1500735

Stark, B. C., & Fukuyama, J. (2020). Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. *Language, Cognition and Neuroscience, 36*(5), 562–585. https://doi.org/10.1080/23273798.2020.1862258

Stark, B. C., & Oeding, G. (2024). Demographic, neuropsychological, and speech variables that impact iconic and supplementary-to-speech gesturing in aphasia. *Gesture, 22*(1), 62–93. https://doi.org/10.1075/gest.23019.sta

Theocharopoulou, F., Cocks, N., Pring, T., & Dipper, L. T. (2015). TOT phenomena: Gesture production in younger and older adults. *Psychology and Aging, 30*(2), 245–252. https://doi.org/10.1037/a0038913

Urena, K., & Stark, B. C. (2024). Characterizing iconic gesture during narratives in chronic traumatic brain injury recovery. *Frontiers in Human Neuroscience, 18,* Article 1393284. https://doi.org/10.3389/fnhum.2024.1393284

van Nispen, K., van de Sandt-Koenderman, M., Sekine, K., Krahmer, E., & Rose, M. L. (2017). Part of the message comes in gesture: How people with aphasia convey information in different gesture types as compared with information in their speech. *Aphasiology, 31*(9), 1078–1103. https://doi.org/10.1080/02687038.2017.1301368

Vanbellingen, T., Kersten, B., Van Hemelrijk, B., Van De Winckel, A., Bertschi, M., Müri, R., De Weerdt, W., & Bohlhalter, S. (2010). Comprehensive assessment of gesture production: A new test of upper limb apraxia (TULIA). *European Journal of Neurology, 17*(1), 59–66. https://doi.org/10.1111/j.1468-1331.2009.02741.x

Walker, G. M., & Schwartz, M. F. (2012). Short-form Philadelphia Naming Test: Rationale and empirical evaluation. *American Journal of Speech-Language Pathology, 21*(2), S140–S153. https://doi.org/10.1044/1058-0360(2012/11-0089)

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition*. Pearson.

Wright, H. H., Koutsoftas, A. D., Capilouto, G. J., & Fergadiotis, G. (2013). Global coherence in younger and older adults: Influence of cognitive processes and discourse type. *Aging, Neuropsychology, and Cognition, 21*(2), 174–196. https://doi.org/10.1080/13825585.2013.794894