Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA)

Manual

Marianne Casilio, MS, CCC-SLP
Manaswita Dutta, PhD, CCC-SLP
Katherine Bryan, MS, CF-SLP
Stephen M Wilson, PhD
Michael de Riesthal, PhD, CCC-SLP

Last Updated: March 2023

The creation of this manual was supported in part by NIH-NIDCD F31DC021108 (PI: Casilio) and NIH-NIDCD R01DC013270 (PI: Wilson). This manual would not be possible without the help of many others who have been involved in the creation of APROCSA throughout the years, including: Pelagie Beeson, PhD, CCC-SLP; Kate Bunton, PhD, CCC-SLP; Kindle Rising, PhD, CCC-SLP; Marja-Liisa Mailend, PhD, CCC-SLP; Audrey Holland, PhD, CCC-SLP; Sarah Schneck, PhD, CCC-SLP; Jillian Entrup, MA, CCC-SLP; Deborah Levy, PhD; Caitlin Onuscheck, MS, CCC-SLP; Lily Walljasper, MA, CCC-SLP; Anna Kasdan, BA; Ellie Tsai, BA; the 2024 MS-SLP graduating class at Vanderbilt University; and all of the individuals with aphasia and student raters who so generously donated their time to participate in our research.

Welcome and thank you for your interest in APROCSA!

We're a group of speech-language pathologists and neuroscientists interested in developing efficient and psychometrically sound assessment tools for clinic and research purposes. The goal of this manual is to create a resource for professionals interested in assessing discourse (also called *connected speech*) using an auditory-perceptual approach. The manual is broken into five primary sections:

1. **Foundational topics: Discourse**
    a. Definition
    b. Theory
    c. Elicitation
    d. Behaviors observed in healthy speakers

2. **Foundational topics: Psychometrics**
    a. Classical test theory, validity, and reliability
    b. Psychometrics of auditory-perceptual assessment

3. **Introducing APROCSA**
    a. Goal and key concepts
    b. Psychometrics of APROCSA

4. **Using APROCSA**
    a. Administration
    b. Rating procedure and scoring
    c. Feature definitions

5. **Resources**
    a. Protocol sheet
    b. References

Our intent is for this manual to serve as a guide while training in APROCSA and when using it after training. There is no need to review the sections in order or in their entirety! Sections 1 and 2 are intended to provide valuable but ultimately optional information on topics relevant to APROCSA. Sections 3, 4, and 5 are intended to supplement formal training in APROCSA and serve as a reference after training is completed.

Through reading this manual, we hope that you, the reader, will be able to:

1. Give a definition of discourse, explain its hypothesized underlying mechanisms and common behaviors, and how it is elicited in a clinical setting.

2. To have an introductory understanding of psychometric principles relevant to measuring discourse and using an auditory-perceptual approach to assessment.

3. Explain the goal of APROCSA and how it is used, as well as briefly summarize its psychometric properties.

4. Describe the administration and scoring procedures, and know how to use the feature definitions and protocol sheet as a future reference.

1. Foundational Topics: Discourse

a. Definition

Discourse refers to the use of spoken and written language in social communication contexts to convey thoughts and meaning. Spoken discourse (also known as connected speech) is commonly defined as beyond the word and utterance level (e.g., Stark et al., 2021) and typically used, as it is this context, to refer to the use of spoken language (as opposed to augmentative-alternative communication or sign).

Spoken discourse is a fundamental aspect of everyday human communication, allowing us to connect with others, convey information, and express our thoughts and feelings in a dynamic and interactive way (Dipper & Pritchard, 2017). It can take different forms: monologic discourse is often used in situations where the speaker presents their ideas, opinions, and information in a structured and organized manner (e.g., speeches, formal presentations). More interactive forms of discourse include language exchange in dialogue or conversations with familiar and unfamiliar communication partners (e.g., casual conversations with friends). Accordingly, language impairments as seen in people with aphasia can significantly influence spoken discourse production and consequently life participation abilities.

According to the International Classification of Functioning, Disability and Health model (World Health Organization, 2018), the assessment and treatment of spoken discourse offers a practical way to address the social participation and activity barriers that individuals with aphasia encounter in daily life due to their communication impairments (Boyle, 2011). In aphasia rehabilitation, researchers and speech-language pathologists have recently shown growing interest in using this approach, as it offers a practical and ecologically valid way to measure the day-to-day challenges faced by individuals with aphasia in social settings due to their communication difficulties (e.g., Stark et al., 2022; Wallace et al., 2017).

b. Theory

Several theories and models have been proposed to improve our understanding of the range of processes involved in discourse processing (e.g., Frederiksen et al., 1990; Kintsch & van Dijk, 1975, 1978). Such frameworks have guided the assessment and treatment of communication impairments in individuals with aphasia.

Frederiksen's three-level framework

Frederiksen et al. (1990) proposed a discourse framework that described three levels of discourse processing occurring sequentially: macrostructural planning, propositional, and linguistic. The *macrostructural planning* stage entails retrieval of a template or schema including rich details about the discourse prompt (e.g., descriptions and facts relating to an event, persons involved, and contextual information). These details are then organized in a logical order for further processing. Next, at the *propositional* level, a set of propositions are generated based on the information retrieved in the macrostructural planning level and arranged using principles of inference and coherence to fit the linguistic frames subsequently. Finally, the language structure is specified at the *linguistic* level wherein sentence frames are generated, syntax is assigned, lexical content is specified, and cohesion is created with the use of reference chains.

Sherratt's multi-level framework

According to the multi-level framework proposed by Sherratt (2007), spoken discourse production involves a number of steps beginning from idea generation to verbal production of utterances. These stages can take place repeatedly and/or simultaneously. As per this framework, discourse processing begins with an input (e.g., tell me about a recent vacation, sharing opinions about current events) which

then facilitates the retrieval of conceptual frames. The type of retrieved frame is specific to the discourse genre/type. Once identified, the overarching discourse structure (e.g., narration, procedural description) guides the content and structuring of the information that needs to be included. The speaker then accesses their semantic and episodic memories to retrieve, synthesize, and integrate the key information with the discourse structure identified earlier. Next, the retrieved information is modified and prioritized based on identification of information that needs to be explicitly produced versus shared knowledge between communication partners. This information is then encoded to generate topic-specific propositions that are sequentially linked together locally to maintain a logical order. Propositions are also grouped semantically and structurally (e.g., number of propositions in a simple sentence). Finally, these temporally sequenced propositions are then linguistically encoded. That is, content words and other lexical and morphological elements are retrieved and specified within the sentence structure. Following the lexical and morphosyntactic processing, the sentences are strung together using cohesive devices/conjunctions and then articulated.

Levelt's discourse model

The discourse model proposed by Levelt (1987; 1990) begins with the retrieval of cognitive (e.g., access to long-term memory) and pragmatic information (new versus shared knowledge) that subsequently go through the *macroplanning* stage where the speaker plans and sequences their generated ideas and monitors their discourse by moving attention from one topic to the other. Next, in the *propositional* level of processing (referred to the 'conceptualizer') micro-planning of the discourse output takes place. In this stage, the lexical items corresponding to the cognitive and pragmatic units are activated and propositional units (or a "pre-verbal message") are created. During the conceptualization process, the propositions are connected to create meaningful connections using principles of local coherence. These linked propositions are then encoded grammatically and produced verbally.

The Linguistic Underpinnings of Narratives in Aphasia (LUNA) framework

Based on the existing discourse theories and frameworks, Dipper et al. (2021) proposed a unified framework to explain the theoretical underpinnings of spoken discourse production more comprehensively. The Linguistic Underpinnings of Narrative in Aphasia (LUNA) framework of spoken discourse involves four components: *pragmatics*, *macrostructural planning*, *propositional*, and *linguistic*. The *pragmatics* component is a critical piece of discourse processing because it guides the construction of spoken discourse. This is where the speaker determines the context, setting (office or at home), familiarity and relationship with communication partners (familiar or unfamiliar partners), and topic familiarity (introduction of new topic versus shared knowledge) which influences the nature of their language output. This component relies on working memory, monitoring of the environment and comprehension of the communication partners, social awareness and understanding, and theory of mind.

In the *macrostructural planning* component, organizational frames are created. Speakers may create new templates or use older ones to structure their discourse and the template may also vary with the discourse genre (e.g., remembering a recent holiday versus engaging in a conversation with a new person). This component draws on a number of different cognitive skills including working memory, planning, monitoring, long term memory stores such as episodic memory, and prospective memory. The *propositional* component is considered the pre-linguistic stage of processing that feeds into the linguistic component. In this stage, the macrostructural elements are organized into microstructural elements. Decisions are made regarding thematic roles (i.e., role of an agent or object in a proposition). The organization and meaningful links between individual propositions are established (local coherence). This component involves taking perspectives, attention, selection, organization, lexical and semantic processing. The final component is the *linguistic* component where linguistic processing occurs, that is, the syntactic structure is constructed, lexical elements are retrieved, phonological units are assembled, and utterances are articulated. The LUNA framework suggests that discourse production is not always sequential, rather the four framework components closely interact and influence each other through bidirectional feedback.

c. Elicitation

There are several relevant factors when considering how to elicit discourse, including what types of techniques (e.g., cueing hierarchies) or stimuli to use. Importantly, these contextual factors can influence the types of behaviors observed and their psychometric properties.

We are fortunate to work with a larger team of speech-language pathologists who have several years of experience eliciting hundreds of language samples from individuals with aphasia across the continuum of care. We asked them to share some advice based on their experiences, which is as follows:

1. Take advantage of the silence – this is critical in giving individuals enough time to communicate.

2. Be prepared with multiple types of stimuli, communication boards, etc. Also always have a phone or tablet on-hand to look up information!

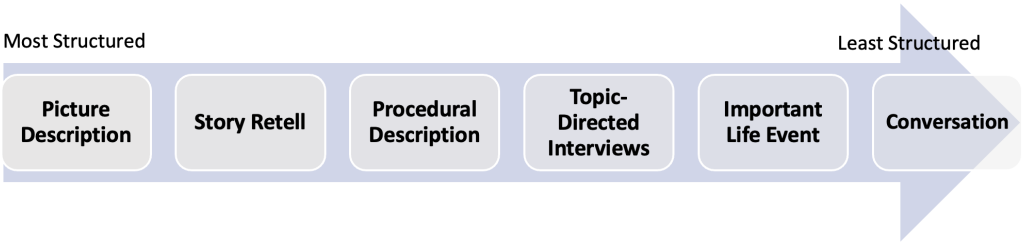
3. Trial a wide range of cues to facilitate success.

4. Balance challenging the person while giving them the opportunity to experience success, too.

Environment
Optimizing the communicative space has long been recognized as an important factor in eliciting language in individuals with aphasia (e.g., Hengst, 2020; Hengst et al., 2019). Ways in which the environment can be optimized include minimizing distractions, placing materials (e.g., pen and paper, pictures) within sight and reach, and building rapport and familiarity between the communication partner and the individual with aphasia. Establishing *common ground* (Wilkes-Gibbs & Clark, 1992)—the creation, maintenance, and recognition of shared histories—is particularly important in eliciting language that is not only rich in output but also improving context so that the communication partner can better judge salient features of speech and language function.

Stimuli and/or prompts
Spoken discourse samples can be collected using a variety of tasks ranging from those that evaluate monologic to interactional aspects of discourse production. Structured and semi-structured prompts that have been commonly employed include single picture (e.g., picnic scene from the Western Aphasia Battery-Revised [WAB-R]; Kertesz, 2007) or picture sequence descriptions (e.g., Broken Window from the AphasiaBank discourse protocol; MacWhinney et al., 2011), procedural descriptions (e.g., describing how to make a peanut butter and jelly sandwich), narratives such as story retelling (e.g., Cinderella story) and free speech samples recounting personal experience (e.g., description of an important life event), topic-directed interviews, role playing, or conversations with familiar and unfamiliar communication partners (see Figure 1). Given that the language output elicited can vary with the discourse stimuli used, collection of multiple discourse samples is recommended to obtain a more comprehensive picture of an individual's language use (e.g., Fergadiotis et al., 2011; Stark, 2019; Wright & Capilouto, 2009).

Figure 1.



d. Behaviors observed in healthy speakers

Normal speech errors

Normal speech errors can take different forms, such as slip of tongue or categorical mis-selections, which can cause a segment to be placed incorrectly within a prosodic "frame" (Berg, 2005; Dell, 1986; Levelt, 1989). Examples of such speech errors may include “new blue blook” for “new blue book” or "a Canadian from Toronto" is mistakenly pronounced as "a Tanadian..." (Fromkin, 1973). These slips can occur as exchange, anticipation, perseveration, or addition errors at different levels such as sound, morpheme, or word level.

Types of speech errors	Examples
Sound level errors	- *Exchange*: "lork yibrary" for "york library" - *Anticipation*: "leading list" for "reading list" - *Perseveration*: "beef needle" for "beef noodle"
Morpheme errors	- *Exchange*: "slicely thinned" for "thinly sliced" - *Anticipation*: "my tow towed" for "my car towed" - *Perseveration*: "rule exsertion" for "explain... rule insertion" - *Addition*: “dedollars deductible” for “dollars deductible”
Word level errors	- *Exchange*: "writing a mother to my letter” for “writing a letter to my mother” - *Anticipation*: “sun is in the sky” for “sky is in the sky” - *Perseveration*: “class will be about discussing the class” for “class will be about discussing the test” - *Addition*: “these purple flowers are purple” for “these flowers are purple”
Note. Examples of errors taken from Stemberger (1982), Fromkin (1971), Garrett (1975), Shattuck-Hufnagel (1979), Dell (1986) and Dell and Reich (1981).	

Normal speech errors in those with and without aphasia

Structural errors can be noted in language output of individuals with agrammatism compared to people with typical language abilities (Stemberger, 1984). Both groups exhibit structural errors, but the nature and frequency of the errors differ: agrammatic speakers tend to make more errors involving the omission of function words and the rearrangement of sentence elements, while the non-agrammatic speakers make more errors involving word substitutions. The structural errors in agrammatic speech may be related to the impairment of specific language processing mechanisms, while those in non-agrammatic speech may reflect a more general failure to access or retrieve the appropriate words.

The types of paraphasias produced by individuals with aphasia versus healthy speakers has also been investigated in picture naming tasks (e.g., Dell et al., 1997). Here, it has been observed that healthy

speakers often produce a small number of paraphasias but they are almost always “mixed” (i.e., have characteristics of both a semantic and phonemic paraphasia, such as saying *rat* for *cat*) or semantically related errors. Individuals with aphasia, however, are much more likely to produce different types of phonemic errors or semantically unrelated errors.

Normal speech errors related to healthy aging

Evidence related to aging-related changes in spoken discourse has been mixed. Older adults can demonstrate difficulties with morpho-syntactic and lexical-semantic processing compared to younger adults despite showing no age-related differences in retrieving phonologically well-formed words. Semantic paraphasias (e.g., “here he’s talking to his mother”, where the speaker intended to say “wife” (Marini et al., 2005 [p. 444])) are common among this age group indicating slowed lexical-semantic processing. Additionally, older adults produce more paragrammatic errors (e.g., errors with use of bound morphemes) suggesting reduced morphological processing. In the context of Levelt’s speech production model (Levelt et al., 1999), access to semantic representations stored in the “lemma” level may deteriorate but phonological encoding at the "lexeme" level is less susceptible to aging and stays relatively preserved across the lifespan.

Age-related declines in syntactic complexity have been documented previously wherein older adults tend to use syntactically less complex structures in their language output (Kemper & Anagnopoulos, 1989; Shadden, 1997). Lexical characteristics may also differ in terms of age: older adults may experience difficulties in the retrieval of pronouns and closed class words (Heller & Dobbs, 1993). In contrast to the abovementioned evidence, Glosser and Deser (1992) found no differences between middle-aged and elderly adult on microlinguistic measures including lexical production errors or syntactic complexity concluding that the microlinguistic structure of discourse is often spared whereas changes occur at the macrolinguistic level. Reduced discourse informativeness and organization are most typically noted (Duong & Ska, 2001; Ehrlich et al., 1997) among older adults. Additionally, global coherence in discourse is influenced by cognitive processes such as working memory and inhibitory control, as well as age. That is, compared to younger adults, older adults exhibit lower levels of global coherence and story construction abilities with missing or ambiguous references (Marini et al., 2005; Wright et al., 2014).

2. Foundational topics: Psychometrics

a. Classical test theory, validity, and reliability

The measurement of discourse is an attempt at quantifying a behavioral phenomenon with a numerical value, which in turn is thought to reflect a psychological construct (Cronbach & Meehl, 1955). In the case of APROCSA, each rating reflects the severity of a person’s performance on a given feature. Collectively, groups of ratings are thought to reflect different constructs, as identified in Casilio et al. (2019): (1) paraphasias, or misselection of words and sounds; (2) logopenia, or paucity of output; (3) agrammatism, or morphosyntactic omissions; and (4) motor speech, or impaired speech motor programming and/or execution.

Importantly and as with all behavioral tests, each rating on APROCSA is an imperfect representation of not only the specific feature but also the targeted psychological constructs. In other words, each rating inherently includes some degree of error. This concept of decomposing an observed test score into its “true” score (i.e., the part of the score that perfectly reflects the psychological construct of interest) and its “error” (.e., the part of that score that reflects the amount of error in its measurement) is called classical test theory (e.g., Crocker & Algina, 1986; Embretson & Reise, 2000):

$$Observed\ Test\ Score\ = True\ Score\ +\ Measurement\ Error$$

As anyone might imagine, the ideal situation is where the observed test score is primarily made up of the true score and not measurement error. In other words, the true score is high and measurement error is low. Quantifying the true score and measurement error is often called “psychometrics,” the properties of which are used to determine whether a test provides valuable information (e.g., in case of the APROCSA, whether the ratings could be used for use in clinical practice or research studies). Psychometrics can broadly be broken into two properties: (1) **validity**, or the extent to which the true score is reflective of the construct of interest; and (2) **reliability**, the degree to which the observed test score contains measurement error and where the measurement error comes from. Importantly, some investigations of validity may be quantitative (e.g., construct or criterion validity, as discussed below) but others may be qualitative in nature (e.g., content validity, as discussed below). Investigations of reliability are always quantitative.

In discourse, relatively little attention has been paid to **validity**, perhaps due to the challenging nature of identifying an underlying theoretical framework (see Section 2 above for greater detail). The majority of studies have focused on correct information units (CIUs; Nicholas & Brookshire, 1994) and on its construct validity (the degree to which a measure aligns with underlying theory),, wherein CIUs have been shown to be highly associated with performance on confrontation naming tests (e.g., Fergadiotis & Wright, 2016; Fergadiotis et al., 2019). Some work has additionally been done regarding the discriminative validity (the degree to which a measure can dissociate groups), showing that select metrics yield distinct scores depending on impairment profile (e.g., Saffran et al., 1989) or group status (healthy versus aphasia; e.g., Nicholas & Brookshire, 1994). The **reliability** of discourse measures has received relatively greater attention. The primary focus has been on inter-rater agreement, or the degree to which two or more examiners agree in identifying and/or coding a given behavior, has been investigated with generally favorable findings (e.g., Rochon et al., 2000; Gordon, 2006). Some emerging work has investigated the contextual reliability of discourse measures, showing that reliability may be contingent on contextual factors like elicitation method (e.g., Stark, 2019)

b. Psychometrics of auditory-perceptual assessment

Most measurement systems for discourse rely on transcription, where audio and/or visual information is first processed into orthography and a coding system before scoring is completed. An alternative approach is to skip the processing step and instead score discourse directly based on the audio and/or visual information. This approach is commonly referred to as “perceptual” and the use of the terms “audio” or “visual” refer to the type of information that is scored. As an example, evaluating discourse and other forms of speech in this manner are commonly referred to as “auditory-perceptual” while one may instead use the term visual when rating features of a videofluoroscopic swallow study.

Auditory-perceptual assessment is a well-established method within special education (e.g., Yoder et al., 2018) and has been extensively applied to the assessment of neurogenic motor speech disorders (e.g., Bunton et al., 2007; Darley et al., 1969a,b; Strand et al., 2014). The Mayo Classification System for Dysarthria (Darley et al., 1969a), a 47-feature system, remains the gold standard approach to the differential diagnosis of dysarthria subtypes. The Mayo system has been shown to not only possess strong inter-rater reliability (Bunton et al., 2007; Darley et al., 1969a) but has multidimensional profiles (Darley et al., 1969a) that are strong associated with distinct etiologies (Darley et al.,1969b). The use of validated auditory perceptual discourse analysis tools in aphasia has been receiving promising and growing support due to its potential to improve time efficiency (Stark et al., 2020).

3. Introducing APROCSA

a. Goal and key concepts

The goal of APROCSA (Casilio et al., 2019) is to help clinicians and researchers obtain a comprehensive diagnostic picture of an individual's strengths and weaknesses on an everyday language skill in an efficient yet precise manner. This is done through rating a comprehensive set of **27 impairment features** that represent breakdown in any or all core speech-language domains (lexical-semantics, morphosyntax, phonology, speech motor programming, speech motor execution).

Assessment approach
APROCSA is an auditory-perceptual diagnostic system, where features or behaviors are judged on the basis of what is heard/perceived. This mirrors the way in which motor speech disorders are most commonly assessed (e.g., Darley et al., 1969a,b).

In its current form, APROCSA focuses almost entirely on the structural or microlinguistic aspects of discourse production. In other words, the focus of APROCSA at the moment is on capturing deficits at the subcomponent level of the linguistic system and understanding how these manifest in a combinatorial way during discourse.

Importantly, APROCSA is an impairment-based diagnostic system in that its features are only intended to capture behaviors that are outside the bounds of healthy speakers. This is in contrast to many popular transcription-based assessment systems for discourse (e.g., CIUs; Nicholas & Brookshire, 1994) on which healthy speakers are expected to exhibit some variability. In other words and from a practical standpoint, it is expected that healthy speakers should receive scores of 0, or Not present (see below for more details on our scoring system) for all or nearly all of the APROCSA features.

Who APROCSA is intended for
At this time, APROCSA has been validated for use by both student clinicians and researchers with experience in aphasia. It is intended for professionals with formal training in linguistics and communication disorders, as many of the features presume requisite knowledge of the domains of language (e.g., phonology, semantics), comorbid disorders following stroke (e.g., apraxia), etc.

Finally, APROCSA currently is only validated for individuals post-stroke. Although APROCSA could be used for other etiologies (e.g., neurodegeneration, TBI, epilepsy) and we certainly welcome further development in this regard, features may behave differently or have psychometric properties that diverge from the ones described here.

b. Psychometrics of APROCSA

The **validity** of APROCSA has been investigated in multiple ways. First and as part of APROCSA's initial development (Casilio et al., 2019), content validity, or the degree to which a test contains all relevant aspects of a construct, was established by conducting a literature review and consulting with other researchers who specialize in discourse. Then, criterion validity, or the degree to which a test aligns with other tests presumed to measure the same targeted construct(s), was investigated by correlating APROCSA ratings on each of the 27 features with (1) analogous features measured in transcription-based discourse (e.g., ratings for the Omission of function words feature were correlated with the number of omissions per hundred words in the available transcripts) and (2) constrained language measures thought to target the same underlying construct (e.g., ratings for the Anomia feature were correlated with the Object Naming Subtest on the WAB-R; Kertesz, 2007). Here, correlations were high (i.e., 24 out of 27 where $r \geq 0.5$), suggesting moderate-to-high criterion validity. Finally, construct validity was investigated by seeing the extent to which the 23 of the 27 features grouped together to reflect the same underlying construct. This was done using a statistical technique

called factor analysis, and four distinct profiles reflecting four constructs were identified, as named above:

1. Paraphasia – misselection of words and sounds
2. Logopenia – paucity of speech
3. Agrammatism – morphosyntactic omissions
4. Motor speech – impaired speech motor processing

Importantly, the patterns among these features were similar to the correlations with constrained language measures, suggesting that the constructs capture in APROCSA reflect relevant speech-language domains (e.g., lexical-semantics, morphosyntax, phonology, speech motor processing). This link between APROCSA and speech-language domains has been further established by investigating its neural correlates, where distinct yet overlapping brain regions are implicated and ones that are strongly associated with specific speech-language domains (Casilio et al., 2022; Casilio et al., under review).

These profiles are used to group and discuss APROCSA's 27 features in detail during our Features Tutorial. We also are in the process of validating a profile scoring system, where select features are rated and then summed to obtain a score for each of the four profiles.

With regard to APROCSA's **reliability**, inter-rater agreement was established by having two groups—12 clinical master's students and 3 researchers—rate 24 samples (Casilio et al., 2019). Here, although researchers were overall more reliable, inter-rater reliability for both groups across the 27 APROCSA features was in the good-to-excellent range. Efforts to evaluate APROCSA's intra-rater agreement, as well as test-retest reliability and contextual reliability, are currently underway.

4. Using APROCSA

a. Administration

A sample of at least five minutes where the individual is speaking is needed to use APROCSA (Casilio et al., 2019), although we are currently piloting whether three minutes is sufficient to observe all relevant behaviors. We measured time by using a stopwatch and total sample time typically ranged between six and seven minutes in length; as such, it is our view that, for most individuals with aphasia, obtaining a sample of a similar duration will likely be sufficient for scoring APROCSA.

A semi-structured interview was used to elicit discourse in our original validation of APROCSA (Casilio et al., 2019). This interview followed the Free Speech protocol developed by AphasiaBank (MacWhinney et al., 2011), where individuals described their speaking abilities, stroke event, and recounted an important life event.

Importantly, the AphasiaBank protocol, as with the protocol used in our research, encourages primarily the use of spoken language to communicate. Use of other modalities, such as writing or gesture, may be helpful in indicating an impairment is present (e.g., a person who finishes an utterance with a gesture as opposed to a spoken word has abandoned that utterance, a behavior we commonly associate with anomia). However, it is important to remember that the majority of features of APROCSA, at least in its current form, focus on spoken language and, as such, spoken language should be encouraged to the extent possible when obtaining a discourse sample.

Given APROCSA's emphasis on spoken language and discourse, not all individuals with aphasia may be able to produce a sample that is long enough to score. In our 2019 paper, we were able to include two individuals with global aphasia and aphasia quotients (AQ) on the WAB-R (Kertesz, 2007) ranged from 20.3 to 92.7. In our more recent work (Casilio et al., 2022; Casilio et al., under review), the overall score on the Quick Aphasia Battery (QAB; Wilson et al., 2018) ranged from 2.42 to 8.81.

Given these findings, it is likely that individuals with a WAB-R AQ of < 20 and a QAB overall of < 2.4 will be unable to produce a sufficient amount of connected speech for APROCSA scoring.

b. Rating procedure and scoring

Rating strategies

As a rater, your job is to listen carefully and determine the appropriate rating for each feature. In order to thoroughly consider each rating scale, the following protocol should be followed when rating each discourse sample:

1. Listen to the sample once. As you listen, rate features as appropriate and take notes on behaviors observed. Please do not pause the video recording. To reduce the cognitive load of judging so many features simultaneously, consider trying the following:

    a. Rate the features that are clearly absent first; alternatively, if a person presents with a particular APROCSA profile (see Features Tutorial for more information), first rate the features most associated with that profile

    b. For features that are clearly present but you're unsure of their severity/frequency, consider writing a ? as a placeholder

    c. Aim to rate at least 50% of the features during the first round if possible

2. Review your scores and notes. Take new notes and/or refer to the descriptions of the discourse features as needed. Make a plan for which features to attend to for the second listen.

3. Listen to the sample again. Verify your ratings and make changes as needed.

Time limits
To minimize listener fatigue, please only listen to a given sample 2 times and spend no more than 15 minutes on a given sample. When rating multiple samples at once, please limit your listening sessions to 1 hour and then take a break before returning for another listening session.

Scoring system
5-point scale, where each point is defined based on both severity and frequency, as adapted from the Apraxia of Speech Rating Scale (Strand et al., 2014):

Not Present (0) = not present or within the range of healthy, older speakers
Mild (1) = Detectable but infrequent
Moderate (2) = Frequently evident but not pervasive
Marked (3) = Moderately severe, pervasive
Severe (4) = Nearly always evident

General scoring considerations
1. Many individuals with aphasia will exhibit only a subset of the features. Moreover, healthy individuals without aphasia will often exhibit some of these features. In particular, healthy speakers commonly retrace, produce false starts, and pause for word finding or other reasons. Some people speak slowly (refer to section 1.d for more information on speech errors in neurologically healthy speakers). It is not uncommon for healthy speakers to produce occasional paragrammatic utterances or to abandon utterances. Consequently, if an individual with aphasia exhibits a feature that would be considered within the expected bounds for a healthy non-elderly person, rate the feature with a score of not present (0).

2. The discourse samples within the APROCSA training may not represent the full spectrum of aphasia severity, particularly those with more severe aphasia. However, APROCSA is designed to capture aphasia severity for all individuals with aphasia who are able to produce at least a few minutes of discourse. Consequently, try to consider the 5-point scale within the context of aphasia severity overall, not simply those with aphasia in these selected speech samples.

3. For individuals with more severe impairments and relatively little output, we recommend having a lower threshold for determining when a behavior is considered impaired versus not. For example, although many healthy individuals may produce occasional semantic paraphasias and the presence of these does not automatically warrant a score of 1 or greater on APROCSA, for an individuals with few utterances, a single semantic paraphasia is likely all that is needed to trigger a score on this feature because, proportional to all of their utterances, this individual is producing this type of error at a higher rate than we would expect in a healthy comparison group.

4. In some forms of aphasia, individuals will attempt to repair their errors. Errors should still be counted as contributing to the relevant feature even if they are successfully repaired. Repairs will generally contribute to the Retracing, False starts, Circumlocution, or Conduite d'approche features.

5. Related to #3, individuals may use a wide range of compensatory strategies to optimize their spoken output. Some examples we have observed include: (1) multimodal communication, (2) generative use of memorized scripts from script training, (3) self-cueing hierarchies (e.g., semantic feature analysis, graphemic-phoneme correspondences); (4) articulatory placement

cues; and/or (5) slowed speech rate or exaggerated prosody. For some individuals, particularly those several years post-stroke, it can be challenging to disentangle whether their behaviors are related to compensation versus impairment. We encourage raters APROCSA to attempt to determine if a behavior is compensatory and, if there is clear evidence that it is, to not rate this as an "impairment" on the features.

6. It's important to balance rating features superficially while also considering underlying theory. For example, an individual may produce telegraphic speech in an effort to be more efficient due to a concomitant motor speech disorder. Superficially, one may want to rate features associated with telegraphic speech (e.g., Omission of function words). However, if there is a clear and consistent pattern suggesting that the behavior is instead linked to an underlying cause that is **not** associated with the speech-language domain of that feature (in this case, morphosyntactic processing), then the feature should not be rated. On the other hand, all behaviors that do not have a clear underlying cause (e.g., ambiguous, multifactorial) should instead be rated superficially; or, in other words, one should attempt to be as objective as possible when judging the presence and severity/frequency of the behavior.

7. Keep in mind that the last feature, Overall communication impairment, is not an average of the other dimensions. In other words, a person does not automatically receive a score of moderate (2) if the majority of the preceding dimensions received a score of moderate (2). The severity of some dimensions (e.g., agrammatism) or the effective use of communication strategies (e.g., circumlocution) may influence the overall presentation. As with the other dimensions, try to objectively rate what is present in the sample.

The importance of utterance segmentation

Reliable rating of many features of APROCSA requires accurate utterance segmentation, a challenging task! Although segmenting discourse into utterances is ultimately somewhat subjective, we recommend using the following as guidelines:

1. A sentence is an utterance

2. Sentences conjoined with a coordinating conjunction (e.g., *and*, *or*, *but*) are separate utterances

3. Rising or falling intonation suggests the end of an utterance

4. Pauses are unreliable markers of utterance boundaries in people with aphasia

Rules for utterance segmentation can also be useful in rating self-corrective features, such as False starts or Retracing (as discussed below). For example, evidence that an utterance was retraced, instead of abandoned, would likely be an even intonation or the immediate production of a new word or phrase without any pausing.

Importantly, unless rating self-corrective behaviors, it's important to consider an individual's **best possible utterance** when rating certain features (e.g., Short and simplified utterances, Paragrammatism). In other words, after factoring out corrective behaviors like Retracings or False starts (as described below), was the utterance complex? Grammatically correct? As noted above, errors that are ultimately self-corrected should still be considered and scored but features of grammatical structure should instead be judged based on the utterance that remains after accounting for various self-corrections.

c. Feature definitions

The features below are shown in their original grouping in the Casilio et al. (2019) study. The Features Tutorial provides an alternative structure by which the features can be rated if one has a hypothesis about which profile or profiles a person is most associated with.

Additionally, transcription is used here throughout to give examples of the various behaviors. For those unfamiliar with common transcription notation, a short key is provided below:

<> Retraced or repeated word/phrase
(.) Short pause
(..) Medium-length pause
(...) Long pause
… Abandoned utterance
- False start

Finally, a few linguistic terms are used to describe the features. A brief description of relevant terms is provided below:

Derivational morpheme	affix that changes the meaning of the meaning of a word
Inflectional morpheme	affix that changes the morphosyntactic structure of a word
Content word	also known as an open class word; nouns, verbs, adjectives, and adverbs are all content words; these words have rich semantic associations
Grammatical word	also known as an closed class word or functor; any type of word other than a content word (e.g., article, preposition, determiner); these words tend to have little or no semantic richness
Broad reference pronoun	pronouns that can take on a wide range of meanings (*this*, *that*)
Supraordinate	larger category a noun belongs to (e.g., dogs are a type of *animal*)

APROCSA features	**Definition and examples**
Anomia	Overall impression of word-finding difficulties. Can be instantiated in many different ways, all of which are also measured as separate features: word-finding pauses before content words, abandoned utterances after failing to retrieve a word, commenting on the inability to retrieve or say words, empty speech, circumlocution, or paraphasias.
Abandoned utterances	Utterances are left incomplete. The speaker may move on to another idea, stop talking, attempt to use another modality (e.g., gesture), or give a vague conclusion to the utterance (e.g., shrug shoulders and say “you know").
Empty speech	Speech that conveys little or no meaning due to lack of specificity. Pronouns and nonspecific words such as thing, stuff, and do are substituted for content words. There are relatively more function words than content words.
Semantic paraphasias	Substitution of content words for related or unrelated content words e.g., cat for dog or candle for dog. Substitution of numbers should not be rated here.
Phonemic paraphasias	Substitution, insertion, deletion, or transposition of a clearly articulated phoneme or syllable. The target should be at least partially recognizable, e.g., papple for apple. The speaker is often aware of the error and will try to correct it, contributing to ratings on False starts, Retracing, and Conduite d'approche. Word forms will primarily be nonwords, though real English word responses may also occur. Errors are more likely to be consonant substitutions, occur on multisyllabic words, and occur at the end of words.

Neologisms	Word forms that are not real English words containing numerous substitutions, insertions, deletions, or transpositions of clearly articulated phonemes. Phoneme selection is so severely disorganized that the intended target is only fleetingly apparent, and the speaker usually doesn't recognize the error.
Jargon	Mostly fluent and prosodically correct but largely meaningless speech that contains paraphasias, neologisms, or unintelligible strings. Productions resemble English syntax and prosody.
Perseverations	Repetition of previously used words or utterances in contexts where they are no longer appropriate.
Stereotypies and automatisms	Commonly used words, neologisms, or rote phrases produced with relative ease and fluency, e.g., *tan*, *I know it*, *dammit*. Typically, stereotypies or automatisms are used to communicate information that otherwise cannot be produced volitionally. Pervasive commentary (e.g., *oh my god*) can be stereotyped or automatic in nature. Communicators (e.g., *yes*, *yeah*) generally should not be considered when scoring this feature unless produced with markedly greater ease than all other productions.
Short and simplified utterances	Utterances are reduced in length or complexity, such as argument structure or number of embeddings. Nonsentence responses (e.g., Did you come with your wife? Yes, or Who did you come with? My wife.) should not be considered.
Omission of bound morphemes	Inflectional (work<u>ed</u>, slow<u>est</u>) or derivational (<u>dis</u>honest, drink<u>able</u>) morphemes are not used when they should be. This includes collapsing an inflected verb to its present tense (e.g., saying *come* for *came*). Omission of these elements generally results in ungrammatical utterances (e.g., I am go to the store) and reduces the length and complexity of utterances. A marked rating (3) should be reserved for speech that exclusively or near-exclusively contains single-word utterances that have bound morphemes. A severe rating (4) should be given for speech that is exclusively or near-exclusively uninflected single-word utterances.
Omission of function words	Function words (determiners, prepositions, pronouns, conjunctions, auxiliaries, etc.) are not used when they should be. Omission of these elements generally results in ungrammatical utterances (e.g., I going to the store). A severe rating (4) should be given for speech that consists exclusively or near-exclusively of single-word utterances.
Paragrammatism	Inappropriate juxtapositions of phrases and misuse of words, including violations of part-of-speech constraints and substitutions of grammatical words and morphemes (e.g., *It's so much wonderful, Makes it hard to speech*).
Pauses between utterances	Unfilled or filled (*um, uh*) pauses that occur between utterances. Pauses between examiner's questions and patient's responses should also be considered. Failure to string together multiple utterances when appropriate can be scored here.
Pauses within utterances	Unfilled or filled (*um, uh*) pauses within utterances, typically before content words. Both prevalence and length of pauses should be taken into account.
Halting and effortful	Speaking is labored and consequently uneven. Intonation, rhythm, or stress may be reduced, absent, or inappropriately placed. The speaker may endorse difficulty with efficiently communicating messages. Those

	with marked (3) or severe (4) ratings will have significant difficulty producing spoken language more generally.
Reduced speech rate	The number of words per minute within utterances is reduced. Speaking slowly and pausing counts toward reduced rate.
Retracing	Sequences of one or more complete words are made redundant by subsequent repetitions, revisions, amendments or elaborations, e.g., <the kite is> (.) the boy is flying the kite.
False starts	Partial words are abandoned after one or two phonemes, e.g., it*'s a ca-cat.* This is a self-corrective behavior that may or may not be trained as a communicative strategy.
Conduite d'approche	At least three successive attempts at producing a target that appears to be known to the speaker but may not be clear to the listener (e.g., *st-stun, start, starling, starting* for *startling*). The target may or may not be achieved. The speaker is aware of their errors. These instances also contribute to ratings for False starts, Retracing, or Phonemic paraphasias.
Target unclear	Features (place, manner, or voice) of phonemes are distorted. The target phoneme(s) may or may not be perceptible. In severe cases, the target word or phrase/utterance may be unknown.
Meaning unclear	It is not clear what the speaker is talking about, or the topic may be clear but what is being said about it is not.
Off-topic	It is clear what the speaker is talking about, but it is not clear how it relates to the context.
Expressive aphasia	Language production is disrupted; the speaker experiences difficulty expressing oneself. Disruptions may occur across any or all language domains (i.e., phonology, morphosyntax, lexical-semantics). Receptive language should not be considered.
Apraxia of speech	Speech is slow and contains distortions or distorted substitutions, more commonly in consonant clusters, at the beginning of words, and in longer words. Syllables are segmented within and across words, and phonemes may be elongated, resulting in atypical prosody.
Dysarthria	Speech is difficult to understand and characterized as slurred, choppy, or mumbled. Errors are consistent and are the result of impaired strength, tone, range of motion, or sequencing. Speech breathing, phonation, resonance, articulation, and prosody may be impaired. =
Overall communication impairment	Overall impression of the extent to which the speaker is impaired in conveying their message. A mild rating (1) should reflect an evident speech-language impairment, but no limitation in discussing all topics. A moderate rating (2) should be used when the speaker can readily communicate about simple, everyday topics, but is limited in discussion of more complex topics. A marked rating (3) should be used when communication about everyday topics is possible with help from the examiner, but the speaker shares the burden of communication. A severe rating (4) should be used when all communication is fragmentary, and the examiner carries the burden of communication. These guidelines, including some of the specific wording, are based on the Boston Diagnostic Aphasia Examination Aphasia Severity Rating Scale.

Additional details about the development and validation of APROCSA is available in our 2019 publication in *American Journal of Speech-Language Pathology*. This paper is freely available for download at https://aphasialab.org/aprocsa/.

d. Pilot features

We are currently trialing the inclusion of three additional features within the APROCSA system. The names and definitions for these features are as follows:

APROCSA features	Definition and examples
Circumlocution	Talking around words or concepts without using the target word itself. Can involve one or more utterances (e.g., he was an actor in the '50s who was in lots of movies) or just a list of words (e.g., actor, '50s, movies). This is a self-corrective behavior that may or may not be trained as a communicative strategy.
Omission of key information	Decreased richness of message content due to lack of contextual information and/or relevant story grammar elements (e.g., characters, setting, problem, solution). Limited development of narratives. A score of 0 (not present) applies in the following instances: (i) the speaker includes rich details about characters and settings; conveys multiple plot points; feelings and/or motivation. A score of 2 may be characterized by: (i) inclusion of basic information (e.g., character(s) and type of setting) but limited depth; (ii) vague details; or (iii) succinct responses or narratives that contain fewer descriptions than in a score of 0 or 1. A score of 4 (severe) applies in the following instances: (i) the speaker rarely includes any elements of story grammar; this typically occurs in cases of severely reduced expressive language.
Multimodal repair of communication breakdowns	Use of modalities other than verbal language (e.g., gesturing, writing, using visual supports, humming, etc.) to convey meaning when communication breakdowns occur in verbal communication. Multimodal strategies may be used to repair communication breakdowns by directly conveying information and/or cueing the individual. Examples of successful repairs: (i) *I have kids…one, two, three, four. No* [holds up three fingers] *kids*; (ii) *I have kids…one, two, three, four. No* [holds up three fingers]*…three kids*; (iii) humming the tune of a song, which the listener correctly identifies; and (iv) *they're in…you know…sunny…* [pulls up Florida on a map to show the listener]. A score of 0 (not present) applies in the following instances: (i) no communication breakdowns occur, so repair strategies are not needed; (ii) all communication breakdowns are repaired verbally, so other modalities are not needed; or (iii) all communication breakdowns are repaired through mixed modalities. A score of 2 may be characterized by: (i) inconsistent attempts to repair communication breakdowns; or (ii) inconsistent success of repairs, e.g., some gestures are clear (showing a number on one's fingers, writing an identifiable target word) while others are vague (pointing in a general direction toward something unknown to the listener; attempting to write or draw something in the air, which is not understood). A score of 4 (severe) applies in the following instances: (i) no attempt to repair spoken communication breakdowns by employing other communication modalities.

5. Resources

a. Protocol sheet

Name/identifier _____ Rater _____ Date _____

Rate connected speech using the following scales:

Not present (0) = not present or within the bounds of healthy, non-elderly speakers
Mild (1) = mild impairment or detectable but infrequent
Moderate (2) = moderate impairment or frequently evident but not pervasive
Marked (3) = moderately severe impairment or pervasive
Severe (4) = severe impairment or nearly always evident

FEATURES	0	1	2	3	4
Anomia	not present	mild	moderate	marked	severe
Abandoned utterances	not present	mild	moderate	marked	severe
Empty speech	not present	mild	moderate	marked	severe
Circumlocution	not present	mild	moderate	marked	severe
Semantic paraphasias	not present	mild	moderate	marked	severe
Phonemic paraphasias	not present	mild	moderate	marked	severe
Neologisms	not present	mild	moderate	marked	severe
Jargon	not present	mild	moderate	marked	severe
Perseverations	not present	mild	moderate	marked	severe
Stereotypies	not present	mild	moderate	marked	severe
Short and simplified utterances	not present	mild	moderate	marked	severe
Omission of bound morphemes	not present	mild	moderate	marked	severe
Omission of function words	not present	mild	moderate	marked	severe
Paragrammatism	not present	mild	moderate	marked	severe
Pauses between utterances	not present	mild	moderate	marked	severe
Pauses within utterances	not present	mild	moderate	marked	severe
Halting and effortful speech production	not present	mild	moderate	marked	severe
Reduced speech rate	not present	mild	moderate	marked	severe
Retracing	not present	mild	moderate	marked	severe
False starts	not present	mild	moderate	marked	severe
Conduite d'approche	not present	mild	moderate	marked	severe
Target unclear	not present	mild	moderate	marked	severe
Meaning unclear	not present	mild	moderate	marked	severe
Off-topic	not present	mild	moderate	marked	severe
Expressive aphasia	not present	mild	moderate	marked	severe
Apraxia of speech	not present	mild	moderate	marked	severe
Dysarthria	not present	mild	moderate	marked	severe
Overall communication impairment	not present	mild	moderate	marked	severe
Omission of key information	not present	mild	moderate	marked	severe
Multimodal repair of communication breakdown	not present	mlld	moderate	marked	severe

b. References

Berg, T. (2006). A structural account of phonological paraphasias. Brain and Language, 96(3), 331-356.

Boyle, M. (2011). Discourse treatment for word retrieval impairment in aphasia: The story so far. *Aphasiology*, *25*(11), 1308-1326.

Bunton, K., Kent, R. D., Duffy, J. R., Rosenbek, J. C., & Kent, J. F. (2007). Listener agreement for auditory-perceptual ratings of dysarthria.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, *52*(4), 281.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Clusters of deviant speech dimensions in the dysarthrias. *Journal of speech and hearing research*, *12*(3), 462-496.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of speech and hearing research*, *12*(2), 246-269.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review*, *93*(3), 283.

Dipper, L., Marshall, J., Boyle, M., Botting, N., Hersh, D., Pritchard, M., & Cruice, M. (2021). Treatment for improving discourse in aphasia: a systematic review and synthesis of the evidence base. *Aphasiology*, *35*(9), 1125-1167.

Dipper, L. T., & Pritchard, M. (2017). Discourse: Assessment and therapy. In *Advances in speech-language pathology*. IntechOpen.

Duong, A., & Ska, B. (2001). Production of narratives: Picture sequence facilitates organizational but not conceptual processing in less educated subjects. *Brain and cognition*, *46*(1-2), 121-124.

Ehrlich, J. S., Obler, L. K., Clark, L., & Rockwood, K. (1997). Discourse ability in the healthy elderly. *Journal of communication disorders, 30*(2), 133-149

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Fergadiotis, G., Hula, W. D., Swiderski, A. M., Lei, C. M., & Kellough, S. (2019). Enhancing the efficiency of confrontation naming assessment for aphasia using computer adaptive testing. *Journal of Speech, Language, and Hearing Research*, *62*(6), 1724-1738.

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, *25*(11), 1414-1430.

Fergadiotis, G., & Wright, H. H. (2016). Modelling confrontation naming and discourse performance in aphasia. *Aphasiology*, *30*(4), 364-380.

Frederiksen, C. H., Bracewell, R. J., Breuleux, A., & Renaud, A. (1990). The cognitive representation and processing of discourse: Function and dysfunction. *Discourse ability and brain damage: Theoretical and empirical perspectives*, 69-110.

Fromkin, V. A. (1973). The non-anomalous nature of anomalous utterances. *Speech errors as linguistic evidence*, 215-242.

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.

Gordon, J. K. (2006). A quantitative production analysis of picture description. *Aphasiology*, *20*(02-04), 188-204.

Heller, R. B., & Dobbs, A. R. (1993). Age differences in word finding in discourse and nondiscourse situations. *Psychology and aging*, *8*(3), 443.

Hengst, J. A. (2020). *Understanding everyday communicative interactions: Introduction to situated discourse analysis for communication sciences and disorders*. Routledge.

Hengst, J. A., Duff, M. C., & Jones, T. A. (2019). Enriching communicative environments: Leveraging advances in neuroplasticity for improving outcomes in neurogenic communication disorders. *American Journal of Speech-Language Pathology*, *28*(1S), 216-229.

Kemper, S., & Anagnopoulos, C. (1989). Language and aging. *Annual review of applied linguistics*, *10*, 37-50.

Kertesz, A. (2007). Western Aphasia Battery--Revised.

Kintsch, W., & Van Dijk, T. A. (1975). Comment on se rappelle et on résume des histoires. *Langages*, (40), 98-116.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological review, 85(5), 363.

Levelt, W. J. M. (1987). The perceptual loop theory of speech production, *Psychological Review, 94* (1), 101-124.

Levelt, W. J. M. (1990). *Speaking: From intention to articulation.* MIT Press. Cambridge, MA.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25* (11), 1286-1307.

Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of psycholinguistic research*, *34*(5), 439-463.

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, Language, and Hearing Research*, *36*(2), 338-350.

Rochon, E., Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (2000). Quantitative analysis of aphasic sentence production: Further development and new data. *Brain and language*, *72*(3), 193-218.

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and language*, *37*(3), 440-479.

Shadden, B. B. (1997). Discourse behaviors in older adults. *Seminars in speech and language*, 18(2), 143-157.

Sherratt, S. (2007). Multi-level discourse analysis: A feasible approach. *Aphasiology*, *21*(3-4), 375-393.

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.

Stark, B. C., Dutta, M., Murray, L. L., Bryant, L., Fromm, D., MacWhinney, B., ... & Sharma, S. (2021). Standardizing assessment of spoken discourse in aphasia: A working group with deliverables. *American journal of speech-language pathology*, *30*(1S), 491-502.

Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., ... & Roberts, A. C. (2022). Spoken discourse assessment and analysis in aphasia: An international survey of current practices. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4366-4389.

Stemberger, J. P. (1984). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology*, *1*(4), 281-313.

Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. (2014). The Apraxia of Speech Rating Scale: A tool for diagnosis and description of apraxia of speech. *Journal of communication disorders*, *51*, 43-50.

Wallace, S., Worrall, L., Rose, T., Le Dorze, G., Cruice, M., Isaksen, J., Kong, A., Simmons-Mackie, N., Scarinci, N. & Gauvreau, C. (2017). Which outcomes are most important to people with aphasia and their families? An international nominal group technique study framed within the ICF. *Disability and Rehabilitation, 39*(14), 1364-1379.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of memory and language*, *31*(2), 183-194.

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PloS one*, *13*(2), e0192773.

Wright, H. H., & Capilouto, G. J. (2009). Manipulating task instructions to change narrative discourse performance. *Aphasiology*, *23*(10), 1295-1308.

Wright, H. H., Koutsoftas, A. D., Capilouto, G. J., & Fergadiotis, G. (2014). Global coherence in younger and older adults: Influence of cognitive processes and discourse type. *Aging, Neuropsychology, and Cognition*, *21*(2), 174-196.

Yoder, P. J., Lloyd, B. P., & Symons, F. J. (2018). *Observational measurement of behavior . Baltimore, MD: Paul H*. Brooks Publishing Company.