# An Auditory-Perceptual Rating of Connected Speech in Aphasia

# AN AUDITORY-PERCEPTUAL RATING OF CONNECTED SPEECH IN APHASIA

by

Marianne Casilio

---

A Thesis Submitted to the Faculty of the

DEPARTMENT OF SPEECH, LANGUAGE, AND HEARING SCIENCES

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2017

## STATEMENT BY AUTHOR

The thesis titled *An Auditory-Perceptual Rating of Connected Speech in Aphasia* prepared by *Marianne Casilio* has been submitted in partial fulfillment of requirements for a master's degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Marianne Casilio

## APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Pélagie Beeson 3/27/2017

*Pélagie Beeson* Date
*Professor and Head of Speech, Language, and Hearing Sciences*

Stephen M. Wilson 3/27/2017

*Stephen M. Wilson* Date
*Assistant Professor of Hearing and Speech Sciences*
*Vanderbilt University*

## Acknowledgements

This research was supported in part by National Institutes of Health (NIDCD R01 DC013270). I thank my thesis committee members—Stephen M. Wilson, Pélagie Beeson, Kate Bunton, and Kindle Rising—for their mentorship throughout the development of this project and completion of this manuscript. I thank the student raters who volunteered their time to evaluate the speech samples. I also thank Marja-Liisa Mailend, Chelsea Bayley, and Audrey Holland for their advice and input in the development of the APROCSA. I additionally thank Davida Fromm for her assistance in utilizing the AphasiaBank database, the individuals with aphasia who consented to share their speech samples through AphasiaBank, and the researchers who made these data available to the community.

# Table of Contents

Abstract…………………………………………………………………………………....5

Introduction………………………………………………………………………………..6

Method……………………………………………………………………………….…9

Results……………………………………………………………………………………17

Discussion………………………………………………………………………………..24

References………………………………………………………………………………..31

Tables…………………………………………………………………………………….34

Figure Captions…………………………………………………………………………..39

Figures……………………………………………………………………………………42

Appendices……………………………………………………………………………….48

## Abstract

*Purpose:* The goal of this study was to develop a novel tool for connected speech analysis in aphasia, so that spoken output can be characterized in a data-driven and explanatory manner.

*Method:* We designed a multidimensional rating scheme called the Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA), in which 27 common features were each rated on a 5-point scale. Three researchers and twelve student clinicians rated 24 connected speech samples from the AphasiaBank database.

*Results:* Ratings conducted by both researchers and student clinicians demonstrated good-to-excellent reliability and strong concurrent validity with AphasiaBank measures derived from transcriptions, clinical measures, and subscores from the Western Aphasia Battery (WAB). Factor analysis revealed that four underlying factors—Paraphasia, Logopenia, Agrammatism, and Motor speech—accounted for 79% of the variance in the connected speech profiles. Examination of individual patient scores showed considerable diversity of factor scores among patients of any given aphasia subtype.

*Conclusions:* The APROCSA proved to be a reliable, valid, and efficient tool for research or clinical purposes. The preliminary findings of the factor analysis suggest a parcellation of non-fluency into three distinct profiles—Logopenia, Agrammatism, and Motor speech—which may occur in conjunction with other non-fluent profiles or with the fluent profile.

## Introduction

Connected speech in individuals with aphasia reflects underlying impairments in any of a number of speech/language domains, including lexical retrieval, phonological encoding, grammatical construction, and articulatory agility. This sensitivity to many different types of disturbances makes connected speech analysis a valuable tool for assessment, diagnosis, and evaluation of treatment outcomes. The goal of this study was to develop a novel tool for connected speech analysis, so that spoken output can be characterized in a data-driven and explanatory manner.

There are two predominant approaches to the analysis of connected speech in aphasia: quantitative linguistic analysis and qualitative rating scales (Prins & Bastiaanse, 2004). Quantitative linguistic analysis (e.g., Saffran, Berndt, & Schwartz, 1989; MacWhinney, Fromm, Forbes, & Holland, 2011) is comprehensive, multidimensional, and largely objective, but is time-consuming and requires highly trained transcribers with substantial knowledge of linguistics and aphasia. Though standardized coding schemes are readily available, the application of such schemes is still ultimately somewhat subjective. For example, one transcriber may judge an utterance as abandoned while another may judge it as retraced. Furthermore, while quantitative linguistic analyses offer the transcriber a wealth of data on discrete behaviors, these data do not always provide an explanatory picture of the patient's deficits. For example, systems such as Codes for Human Analysis of Transcripts (CHAT) do not accommodate coding for distorted phonemic substitutions (MacWhinney, 2000). The transcriber may choose to code them as either a phonological error or, if phonemes are truly indiscernible, as an unintelligible word. As a result, differentiation between phonological errors and motor speech errors is not immediately clear. The transcriber must examine the transcription data for evidence in support of either deficit, such as the number of pausing codes or the presence of neologisms. In instances such as this, the quantitative nature of the analysis itself may preclude the transcriber from readily deriving meaningful information.

In contrast, qualitative rating scales, such as the Western Aphasia Battery (WAB; Kertesz, 1982) fluency rating or the Boston Diagnostic Aphasia Evaluation (BDAE; Goodglass, Kaplan, & Barresi, 2001) profile of speech characteristics, are quick tools intended for use by clinicians. Easy to administer and score, they provide an overall profile of the patient's speech. The design of these instruments, however, presupposes which features are important. For example, the grammatical form feature on the BDAE profile of speech characteristics is defined as the patients' use of morphemes and varied grammatical structures. The rating scale scores, however, are expressed on a continuum of agrammatism, with a score of 1 encoding the absence of syntax, a score of 4 corresponding to simplified structures with omission of morphemes, and a score of 7 used for normal syntax with varied structures. The scale, designed to provide a subtype profile, does not allow for rating of paragrammatism, a well-documented phenomenon in aphasia. Emphasis on subtypes limits the generality of these tools because relevant behaviors may not be captured or appropriately categorized, as the majority of patients do not fit cleanly into a classical aphasia profile (Prins, Snow, & Wagenaar, 1978; Albert et al., 1981). Another disadvantage to qualitative rating scales is that they involve the rating of only one or a few features, and so are not comprehensive. For instance, the WAB fluency rating requires the examiner to consider multiple linguistic domains on the same scale. Similarly, the BDAE profile of speech characteristics includes one scale for paraphasias, regardless of whether they are the result of phonological or semantic deficits. Further limitations include the design of the scales themselves, such as the non-linear scale in the WAB and the inconsistent quantification of scale points in the BDAE, where only the extremes (1 and 7) and the middle (4) are defined.

In this study, we took a different approach to the quantification of connected speech characteristics, inspired by the auditory-perceptual approach to assessment of motor speech disorders (Darley, Aronson, & Brown, 1969a,b, 1975), in which speech samples are rated on a large number of perceptual dimensions. The auditory-perceptual approach is reliable in both experienced and inexperienced listeners (Bunton, Kent, Duffy, Rosenbek, & Kent, 2007), and patterns are associated with distinct etiologies (Darley, Aronson, & Brown, 1969b, 1975). Consequently, this approach remains the gold standard for assessment, diagnosis, and clinical decision-making in motor speech disorders (Duffy, 2013).

We designed a multidimensional Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA) in which 27 features of connected speech were each scored on a 5-point scale. In order to assess the reliability and validity of each feature, connected speech samples from 24 individuals with aphasia were retrieved from the AphasiaBank database (MacWhinney, Fromm, & Holland, 2011) and evaluated by experienced researchers and student clinicians. The data were then examined to examine the reliability and validity of the tool.

We had three aims: (1) to quantify the reliability of the APROCSA in experienced researchers and student clinicians, reflecting two possible ways in which the APROCSA might be used in practice; (2) to assess the concurrent validity of the APROCSA, by examining correlations between APROCSA features and measures derived from quantitative linguistic analysis and established diagnostic measures; and (3) to explore empirically motivated and explanatory underlying factors that explain the patterns among the APROCSA features.

# Method

## Rating scale

Twenty-seven common features of connected speech in aphasia were selected for inclusion in the APROCSA (Table 1). One additional feature, *circumlocution*, was also rated but was subsequently excluded from the analyses due to notably poor inter-rater reliability. Features were grouped into seven categories—lexical retrieval, selection of words and sounds, grammatical construction, rate and timing, self-correction, clarity, and diagnostic—that were identified as representative of the features collectively. Most features were associated with language processing, with motor speech deficits captured broadly with ratings of dysarthria and apraxia of speech. Some features were reflective of both motor speech and language processing (e.g., *halting and effortful*).

Features were identified based on previous methodologies developed for quantitative linguistic analysis (Saffran et al., 1989; MacWhinney, 2000; Wilson et al., 2010; Yagata et al., 2017; McCarron et al., 2017). Each was selected based on the following criteria: (1) its prevalence in speakers with aphasia, as identified by normative data from prior language batteries (e.g., WAB, BDAE); (2) its salience to a listener in the absence of transcribed data; and (3) its ability to reflect deficits in one or more language domains, such as *anomia*, a composite feature designed to capture deficits across lexical access, phonology, and semantics.

The features of the APROCSA were defined superficially, requiring the rater to only consider what they hear, rather than attempt to identify which feature(s) are associated with different language processes. For instance, the feature *short and simplified utterances* may be a derivative of grammatical and/or motor speech deficits. However, raters were explicitly instructed to not consider the underlying impairment when rating the feature. In other words, rating of the feature did not rely on *a priori* knowledge of aphasia typology or models of language processing.

A 5-point, equally-appearing interval scale was used to rate each feature (Table 2; Strand, Duffy, Clark, & Josephs, 2014; Bunton et al., 2007). Each point on the scale was explicitly defined, accounting for both severity and frequency. Importantly, a score of 0 was defined as being within the expected bounds of healthy, non-elderly adults. Individuals without aphasia may occasionally exhibit some of the features identified in the APROCSA, such as retracing a phrase or pausing for word-finding or other reasons. Similarly, individuals with aphasia may only present with a subset of the defined features of the APROCSA.

The APROCSA was designed to be an efficient tool that could be completed by an experienced clinician or researcher in approximately five to ten minutes. The resultant product was a one-page score sheet that consisted of the 5-point scale definitions and a list of all 27 features (Appendix 1). A 4-page manual with general administration considerations and brief explanations of each connected speech feature was also created to accompany the score sheet (Appendix 2).

### Connected speech samples

Twenty-four videotaped connected speech samples of speakers with chronic post-stroke aphasia (aged 49 to 76 years, 12 males) were selected from the AphasiaBank database (MacWhinney, Fromm, & Holland, 2011). All speakers were right-handed monolingual English speakers with vision and hearing (aided or unaided) adequate for testing. Demographic information and standardized test scores are presented in Table 3.

Samples were collected at participating universities and outpatient clinics across the country. The samples were selected such that patients were diverse in aphasia severity (Aphasia Quotient (AQ) range 20.3 to 92.7) and subtype (7 Anomic, 5 Conduction, 4 Wernicke’s, 4 Broca’s, 2 Global, 1 Transcortical Motor, 1 Transcortical Sensory). WAB subtype ratios were intended to approximately reflect prevalence of subtypes within typical outpatient populations of individuals with aphasia (Kertesz, 1979). Furthermore, within each WAB subtype, patients were strategically selected at equally appearing intervals to represent a range of AQ severity.

Excerpts were clipped to approximately 5 minutes to broadly capture all connected speech features identified on the APROCSA, as previous research identified this time frame as adequate to evaluate communicative efficacy in aphasia, assuming all diagnostic behaviors occur at least three times per minute (Boles & Bombard, 1998). All excerpts were selected from the *Free Speech Samples* portion of the AphasiaBank protocol, during which patients talked about their speaking abilities, stroke, recovery, and in some cases recounted a memorable life event. Samples with less than five minutes of recorded speech in these areas were not considered for inclusion.

## Raters

Two groups of raters were included in the study. The first group included three expert researchers with experience in the analysis of connected speech (SMW, KR, MC). SMW was an aphasia researcher with 14 years of experience in aphasia research and extensive experience with connected speech analysis. KR was a licensed speech-language pathologist with more than 10 years of experience as a research clinician in an aphasia research laboratory. MC was a clinical master’s student at the University of Arizona with 3 years of experience in transcription and specific training in the transcription of connected speech in aphasia. The second group was comprised of 12 clinical master's students at the University of Arizona with academic and clinical training in aphasia (Table 4). Raters in both groups passed a hearing screening and spoke English with native proficiency.

The study was approved by the University of Arizona Institutional Review Board. Student raters provided written informed consent for the study and were modestly compensated for their participation.

## Rating Procedures

**Expert calibration.** Prior to rating speech samples, one sample from AphasiaBank, elman03a, was selected for rating calibration and discussion among the expert raters. Elman03a was a 52-year-old male who was 11-years post-stroke. His WAB AQ was 66.2 with a Broca's subtype. He also had a clinical diagnosis of apraxia of speech. Inclusionary criteria for this sample was the same as listed above except for language status, as elman03a spoke English and Mandarin. This particular speech sample was selected because elman03a was one of the few speakers with relatively moderate aphasia who presented with almost all of the APROCSA features.

The three expert raters evaluated elman03a independently and then met to discuss their ratings. For any feature that did not demonstrate exact agreement, a consensus score was reached through discussion and re-watching the videotaped sample. The videotaped sample and consensus scoring was then used as part of the training session developed for student raters, which is discussed below.

**Expert rating procedures.** The expert raters then evaluated all 24 patient samples. They were instructed to watch each patient sample and rate all features simultaneously. The 4-page manual was provided as a reference. Restrictions regarding rating time duration were not rigidly enforced, though expert raters were asked to spend no more than 15 minutes on a patient sample. Ratings were completed within a one-month time frame using printouts of the score sheet and a pencil. Videotaped speech samples were viewed using personal headphones and computers. SMW and KR listened to each sample approximately 1.5 times while MC listened to each sample approximately 2.5 times.

**Student training.** Prior to rating speech samples, student raters participated in a 2.5-hour training session that reviewed the purpose of the APROCSA, administration and scoring, and an in-depth explanation of the 27 connected speech features. Trainings were offered on two different dates to accommodate raters' schedules. MC delivered the training presentation with the help of a doctoral candidate at the University of Arizona who led a 20-minute section on differential diagnosis of apraxia of speech, her expertise. The training presentation was followed by a practice session where students rated elman03a independently, reviewed the consensus scores, and discussed any discrepancies in scoring. Student rater questions included clarification on particular APROCSA features, such as differentiation of paragrammatism from agrammatism, and phonological paraphasias from apraxia of speech.

**Student rating procedures.** Each student then rated a randomized selection of 8 of the 24 samples. Randomization was designed to ensure that each of the 24 samples was rated 4 times. Students were instructed to watch each patient sample twice and rate all features simultaneously, spending no more than 15 minutes per patient sample. A brief break between the first and second listen was permitted to review notes and scores. Sessions consisted of two appointments over a 2-week period, during which ratings were completed using printouts of the score sheet and a pencil. Videotaped speech samples were viewed using a Lenovo ThinkPad T60 laptop and Audio-Technica QuietPoint ATH-ANC7b headphones. Appointments were limited to 1 hour in length to control for fatigue (Bunton et al., 2007). As with the expert raters, students were given the 4-page manual as a resource.

**Reliability**

The reliability of each feature was assessed in terms of intraclass correlation coefficients (ICCs) using models described in McGraw & Wong (1996). For expert raters, we calculated ICCs for two-way models, as each of the 24 patients was rated by all 3 of the experts. Both the patients rated and the expert raters were treated as random factors (i.e., experts were in principle drawn from a pool of experts), though it is important to note that there is no difference in the calculation of ICCs whether experts were considered random or fixed. Absolute agreement, as opposed to consistency, was identified as an area of interest, so that systematic differences between experts regarding whether they assigned relatively high or low scores to a particular variable would be reflected in a reduction in the estimate of reliability. As such, the appropriate ICCs for the expert group were ICC(A,1), which estimates the absolute agreement of any two measurements, and ICC(A,k), which estimates the absolute agreement of measurements that are averages of *k* independent measurements, where *k* = 3 (because three experts rated each patient). In other words, ICC(A,1) is an estimate of reliability in a situation where patients are rated by a single expert, whereas ICC(A,k) is an estimate of reliability in a situation where patients are rated by averaging the ratings of 3 experts.

For students, we calculated ICCs for a one-way model in which patients rated were a random factor. Each patient was rated by 4 students, but because a different subset of students rated each patient, there was no inherent order to the 4 ratings obtained for each patient. As a result, the appropriate ICCs in this situation were ICC(1), which estimates the absolute agreement of any two measurements, and ICC(k), which estimates the absolute agreement of measurements that are averages of *k* independent measurements, where *k* = 4 (because four students rated each patient). In other words, ICC(1) corresponds to an estimate of reliability if patients were rated by single random students drawn from the population of students we have described, whereas ICC(k) corresponds to an estimate of reliability in the situation where patients were rated by averaging the ratings of 4 students drawn from this population.

The reliability of each individual expert and each individual student on each APROCSA feature was assessed by calculating an ICC (type A,1) between the individual and the mean of the other two experts (in the case of experts), or the mean of the three experts (in the case of students), on the relevant set of rated patients (24 for experts, 8 for students). For each individual, the 27 ICCs (one per variable) were converted to *z*-scores (McGraw and Wong, 1996, Appendix B), averaged together, and converted back to *r*. The mean ICCs of the experts and students were then compared with a 2-sample *t*-test (equal variance, one-tailed).

## Validity

The concurrent validity of the APROCSA connected speech features, based on the mean of the 3 experts, was investigated by calculating Pearson correlations with 25 AphasiaBank measures, including 17 quantitative linguistic measures, two motor speech measures (clinical diagnoses of apraxia and dysarthria), and six WAB measures (AQ and subscores for information content, fluency, comprehension, repetition, and naming). Quantitative linguistic measures were derived from transcriptions coded by AphasiaBank administrators using CHAT and calculated using FREQ and EVAL analyses in Computerized Language ANalysis (CLAN; MacWhinney, 2000).

FREQ analysis performs frequency counts of designated word-level and utterance-level error codes, such as morphosyntactic errors, and post codes, which capture other utterance-level phenomena, such as retracing or pausing. EVAL analysis performs calculations that are commonly used by aphasia researchers or clinicians, such as mean length of utterance (MLU). The majority of measures were presented as proportions, either in comparison of two part-of-speech elements, such as pronouns and nouns, or per hundred words (phw). A description of each measure, along with its relevant CHAT code(s) is provided in Table 5.

Twenty-four of the 27 APROCSA connected speech features were identified *a priori* as representing a similar construct to one or more AphasiaBank measures. Importantly, neither the selected AphasiaBank measures nor the APROCSA features was chosen with the goal of duplicating the other. Rather, the measures were selected in an effort to identify and analyze existing correspondences.

The remaining three APROCSA features—*paragrammatism, perseverations*, and *stereotypies*—were initially identified *a priori* as having related AphasiaBank measures but were not included in our analysis due to insufficient use of their corresponding CHAT code(s). In the case of *paragrammatism*, the CHAT manual defines several word-level codes designed to capture paragrammatism; however, only one of these codes was used in the transcripts we reviewed. The code [+gram], a measure of ungrammatical utterances, was also designed for coding of paragrammatism, as well as agrammatism, though this code appeared to be used primarily for agrammatic utterances in our transcripts. Similarly, the CHAT codes identified as representative of the *perseverations* and *stereotypies* features, [+per] and [*n:uk:s] respectively, were not present in our patient transcripts and consequently could not be analyzed.

### Patterns

To examine relationships among the APROCSA features, pairwise Pearson correlations were first computed between all features.

Then, factor analysis with varimax rotation was performed based on the mean of the 3 experts. Four connected speech features—*conduite d'approche*, *off-topic*, *dysarthria*, and *overall communication impairment*—were removed from analysis, as the algorithm required fewer features than patients. Three of the four features—*conduite d'approche*, *off-topic*, and *dysarthria*—were excluded due to their relatively low reliability and relatively restricted distribution among the patient samples. *Overall communication impairment* was removed to decrease redundancy in the analysis, as it was similar to and highly correlated with the *expressive aphasia* feature.

Factor analysis was performed using *factoran* in MATLAB (Mathworks, Natick, MA). Four factors were found to be the most meaningful reduction of the data, as described in the results section. Correlations between the resultant factors and AphasiaBank measures were calculated using pairwise Pearson correlations. Finally, the factor loadings for each patient were derived from the results of the factor analysis.

## Results

Most APROCSA features demonstrated wide distributions among the 24 patients (Figure 1) for both the expert and student raters, showing that the selected patient sample varied in terms of its presenting features and the severity of those features.

## Reliability

ICCs of type (A,k), an estimate of reliability when ratings were averaged across 3 experts, were excellent ($r \geq 0.75$) for 19 features, good ($0.60 \leq r < 0.75$) for 6 features, and fair ($0.40 \leq r < 0.60$) for 2 features (Figure 2). ICCs of type (A,1), an estimate of reliability in a situation where patients were rated by a single expert, were excellent for 7 features, good for 6 features, fair for 10 features ($0.40 \leq r < 0.60$), and poor for 3 features ($r \leq 0.40$).

Results for ICC(k), an estimate of reliability where patient scores were averaged across 4 students drawn from the population of students described, demonstrated excellent reliability for 11 features, good for 12 features, fair for 2 features, and poor for 2 features. ICC(1), an estimate of reliability where patients were rated by single random students drawn from the population of students described, showed excellent reliability for 4 features, good for 7 features, fair for 12 features, and poor for 3 features.

The mean ICCs of the three experts were very similar (SMW: 0.68; MC: 0.69; KR: 0.69), while the students were much more variable (mean = 0.56 ± 0.11 SD, range 0.42 to 0.70). As a group, the experts were more reliable than the students ($t(13) = 1.90$, $p = 0.040$), but at least 3 of the 12 students were as reliable as the experts (means of 0.68, 0.70 and 0.70), suggesting that a subset of students can be identified who will perform comparably to experts. Given that the expert group was found to be more reliable, subsequent assessment of validity and factor analyses were calculated using the mean of the expert raters' scores.

## Validity

Concurrent validity was assessed by examining correlations between each of the 27 APROCSA feature and the 25 measures selected from AphasiaBank (Figure 2). As mentioned above, 24 out of 27 APROCSA features were identified *a priori* as representing a similar construct of one or more AphasiaBank measures, which are outlined in yellow in Figure 2. Of the 24 APROCSA features examined, 18 showed strong correlation(s) (|*r*| ≥ 0.5) with at least one of the relevant measure(s). For example, correlations between the APROCSA feature *omission of function words* and the CHAT transcript measures closed class words (proportion) and agrammatic utterances (phw) were -0.70 and 0.90 respectively. Two of the 24 features—*off-topic* and *dysarthria*—demonstrated significant but not strong correlations with their corresponding measure(s). Four of the 24 features—*semantic paraphasias, phonemic paraphasias*, *conduite d'approche*, and *apraxia of speech*—did not exhibit a significant correlation with their respective AphasiaBank measure(s). Despite the lack of correlation with the aforementioned features, the correlations observed for the great majority of features support the validity of the APROCSA, and many of the failings of the correlations were likely due to inherent limitations of the AphasiaBank measures, as explained in the discussion.

## Patterns

**Correlations among APROCSA features.** Pearson correlations between each pair of APROCSA features were computed (Figure 3). As anticipated, there were many instances in which pairs of APROCSA features correlated strongly (|*r*| ≥ 0.5) with one another. For example, the correlation between *omission of bound morphemes* and *omission of function words* was 0.92. In some instances, features were anticorrelated, such as a correlation of -0.45 between *semantic paraphasias* and *omission of bound morphemes*. Given these findings, factor analysis was performed to further define the relationships among the APROCSA features.

**Factor analysis.** Patterns among the APROCSA features were identified using factor analysis (Figure 4). A model with four factors proved to provide the most explanatory dimensionality reduction of the data, accounting for 79.5% of the variance in the data. We labeled the factors Paraphasia, Logopenia (paucity of speech), Agrammatism, and Motor speech, based on the features that loaded on them, as described in detail below. The eigenvalues of these factors were 5.31, 5.21, 4.38 and 3.39, and the percentage of variance explained was 23.1%, 22.6%, 19.1% and 14.7% respectively. Communality values of the APROCSA features ranged from 0.56 to 0.97, indicating that a high proportion of the variance for each feature was explained within the four factors.

Models with fewer than four factors conflated one or more of these four factors, and explained substantially less of the variance in the data. In particular, a two-factor model conflated the Logopenia, Agrammatism, and Motor speech factors, and explained only 58.8% of the variance, whereas a three-factor model conflated the Logopenia and Motor speech factors, and explained only 70.2% of the variance. In contrast, a five-factor model yielded four factors similar to those identified in the four-factor model, as well as an additional factor with an eigenvalue of 0.73 (i.e., less than 1) that explained only 3.2% of the variance, and the factor loadings of which had no obviously meaningful interpretation.

The four factors were representative of a constellation of phenomena associated with fluent (Paraphasia) or non-fluent (Logopenia, Agrammatism, Motor Speech) aphasia profiles. The Paraphasia factor was characterized by paragrammatic utterances that frequently contained selection errors in phonology and semantics, with heavy factor loadings on the *paragrammatism*, *semantic paraphasias*, *phonemic paraphasias*, *neologisms*, and *jargon* features. Other characteristic features included empty utterances that were abandoned and retraced, as evidenced by loadings on *empty speech*, *false starts*, *abandoned utterances*, and *meaning unclear*.

The Logopenia (paucity of speech) factor represented patients with significant anomia who produced halting and slow speech punctuated by frequent pausing and perseverations, as represented by positive loadings on the *anomia*, *halting and effortful*, *pauses between utterances*, *pauses within utterances*, *reduced speech rate,* and *perseverations* features. Furthermore, utterances were often short and abandoned with a poorly understood message, as evidenced by loadings on the *short and simplified utterances, meaning unclear,* and *abandoned utterances* features. Notably, phenomena associated with grammatical form were not characteristic of the Logopenia factor, with minimal loadings on the *omission of function words* and *omission of bound morphemes* features, and a negative loading on the *paragrammatism* feature.

In contrast, the Agrammatic factor was characterized by simplified, unclear utterances with frequent stereotyped phrases and grammatical omissions, as evidenced by heavy loadings on the *short and simplified utterances*, *meaning unclear*, *stereotypies*, *omission of function words*, and *omission of function words* features. Retracing and false starts were infrequent, with negative loadings on the *retracing* and *false starts* features.

The Motor speech factor was representative of patients whose speech was halting, effortful, slow, and contained frequent pausing, as evidenced by heavy loadings on the *halting and effortful*, *slow speech rate*, *pausing between utterances*, and *pausing within utterances* features. Phonemes were distorted or imprecise and motor planning deficits were evident, with positive loadings on the *target unclear* and *apraxia of speech* features.

**Factor analysis correlations with AphasiaBank measures.** Another way of understanding the meaning of the four factors was to correlate them with the 25 previously examined AphasiaBank measures (Figure 5). Data from this analysis provided a similar picture to the previous correlation analysis of APROCSA features. Patterning of correlations in the quantitative transcription, WAB, and motor speech measures from AphasiaBank appeared congruent with the factors identified with the APROCSA.

To determine whether similar factors would emerge from quantitative linguistic measures from AphasiaBank, factor analyses were run on these 17 measures. Models with between 2 and 7 factors explained between 66.6% and 90.8% of the variance with all factors having eigenvalues greater than 1 and explaining non-trivial proportions of the variance. However, the factors tended to have much less clear interpretations than those derived from the APROCSA variables. For example, the 4-factor analysis explained 80.0% of the variance, with three factors seeming to reflect Agrammatism, Empty speech, and Paraphasia (but not phonemic paraphasias), and a fourth factor that loaded heavily on Phonemic paraphasias but the other loadings of which were difficult to interpret. In short, factor analyses based on quantitative linguistic measures from AphasiaBank yielded factors that were only sometimes readily interpretable in terms of underlying deficits.

**Factor loadings by patient.** The factor loadings for individual patients were plotted and showed considerable diversity among patients of any given aphasia subtype (Figure 6). For instance, patients with nonfluent aphasia subtypes loaded on several of the nonfluent factors. The majority of those with Broca’s aphasia loaded on the Agrammatism factor (scale33a, TCU08a, BU08a), as expected, though one patient (BU08a) loaded on the Motor speech factor and three patients (TCU08a, TAP11a, BU08a) loaded on the Logopenia factor to varying degrees. The two patients with Global aphasia presented with remarkably different profiles, one who loaded moderately on Logopenia and Agrammatism (scale09a) and the other who loaded heavily on Paraphasia (TAP09a).

Significant variety was also observed among the patients identified with fluent aphasias, with loadings observed on both fluent and non-fluent factors. Of the four with Wernicke's aphasia, only one loaded on the Paraphasia factor (elman14a). One (kurland18a) loaded heavily on the Logopenia factor, while another (thompson05a) loaded on the Agrammatism factor. The fourth patient (elman12a) had no positive loadings. Of those with Conduction aphasia, two of the five (kurland20a, TCU07a) demonstrated loadings on the Paraphasia factor, with one (TCU07a) loading additionally on the Logopenia factor. Two additional patients (willamsom04a, ACWT09a) loaded primarily on the Logopenia factor. One (wright203a) had no positive loadings. Of those with Anomic aphasia, one patient (adler01a) loaded heavily on the Motor speech factor, while the rest demonstrated relatively small loadings across the Motor speech (TAP18a, whiteside06a), Paraphasia (adler01a, kurland07a), and Agrammatism (scale30a) factors. Two (fridriksson05a, kurland28a) appeared to have no positive loadings.

In looking at the factors, patients with similar qualities in their connected speech were identified as having a variety of different aphasia subtypes. For instance, patients who loaded on the Paraphasia factor spanned a wide range of subtypes and AQ severities (86.8 to 20.5). Four of the 8 patients who loaded on the Paraphasia factor had Anomic aphasia (kurland20a, TCU07a) or Conduction aphasia (adler01a, kurland07a). The remaining patients either had Wernicke's aphasia (elman14a), Broca's aphasia (BU08a), Global aphasia (TAP09a), or Transcortical sensory aphasia (williamnson16a). Similarly, AQ severities were varied for those who loaded on the Agrammatism factor (90.3 to 20.3). Five of the 10 patients who loaded on this factor had Broca's aphasia (scale33a, TCU08a, BU08a) or Global aphasia (TAP09a, scale09a). Two patients each had Anomic aphasia (TAP18a, scale30a) and Conduction aphasia (ACWT09a, williamson04a). One also had Wernicke's aphasia (thompson05a).

Those who loaded on the Logopenia factor were primarily individuals with AQ severities in the 50s or lower (58.1 to 20.3). Five of the seven patients had Broca's aphasia (TCU08a, TAP11a, BU08a) or Global aphasia (TAP09a, scale09a). The remaining two had Wernicke's aphasia (kurland18a) or Conduction aphasia (TCU07a). Patients with loadings on the Motor speech factor were highly diverse in their AQ severity range (90.3 to 20.5) but primarily loaded on Anomic aphasia (TAP18a, whiteside06a, adler01a) and nonfluent aphasias associated with co-morbid motor speech disorders, such as Transcortical motor aphasia (ACWT02a), Broca's aphasia (scale33a, BU08a), and Global aphasia (TAP09a).

Of note, 4 of the 24 patients did not load positively on any factor. All had AQ severities in the 70s or higher (74.4 to 92.7) and had a fluent aphasia subtype: two had Anomic aphasia (kurland28a, fridriksson05a); one had Conduction aphasia (wright203a); and one had Wernicke's aphasia (elman12a).

## Discussion

The APROCSA proved to be an efficient, reliable, and valid means of characterizing connected speech in aphasia, revealing explanatory factors underlying the multidimensional profiles observed. It warrants further investigation as a tool for assessment, diagnosis, and evaluation of treatment outcomes in aphasia.

### Reliability

The APROCSA was observed to be a reliable tool for quantifying connected speech in aphasia, with raters from both groups demonstrating good-to-excellent reliability on the majority of the features. While experienced researchers were generally more reliable than student clinicians, a subset of student clinicians performed comparably to researchers, suggesting that extensive experience is not prerequisite to being a good rater. Given the reliability demonstrated by both groups, the APROCSA shows potential for use as an assessment tool in research and clinical settings. Experienced speech-language pathologists or aphasia researchers may use the APROCSA to efficiently and reliably capture characteristics of connected speech in aphasia. The data-driven approach to the APROCSA, coupled with its relatively simple administration and rating scheme, makes it an attractive tool for aphasia assessment. Student research assistants may also serve as effective raters, though the increased variability observed within the student rater group suggests that structured training and screening of students by comparing performance to the data described here may be necessary for identifying reliable raters.

Some APROCSA features were shown to be more reliable than others. The lower reliability observed across both groups for *off-topic* may have been due to the relative difficulty in judging the presence and severity of this feature, which requires the rater to make inferences on the quality of the patient’s utterances in response to an examiner’s question. In this regard, *off-topic* requires greater context than the other features, which may have contributed to the variance in rater scores. Inherent characteristics of the feature may have played a role in the relatively low reliability of *phonemic paraphasias*, which captured errors at the level of the phoneme. Errors such as these are often difficult to parcellate from similar sounding errors due to apraxia of speech, such as phonemic distortions. On the other hand, features representative of phonemic paraphasias on larger linguistic units, such as *neologisms* and *jargon*, demonstrated good-to-excellent reliability and correlated strongly with *phonemic paraphasias*. These results suggest that phoneme-level errors were captured by APROCSA features with similar constructs despite the relatively low reliability of the feature itself.

Reliability of the APROCSA was comparable to established assessment measures for aphasia and motor speech disorders. For instance, interjudge reliability of quantitative linguistic analysis has been established in a study examining the Quantitative Phrase Analysis method (QPA; Rochon, Saffran, Sloan, Berndt, & Schwartz, 2000). Reliability for rating agrammatic speakers was determined by comparing scores of two independent raters from randomly selected patient transcripts. ICCs for twelve QPA measures were derived, all of which were in the excellent range, varying from 0.89 (number of embeddings) to 0.98 (number of closed class words, number of pronouns, elaboration of auxiliaries, determiner/noun ratio).

Inter-rater reliability is also commonly reported for qualitative rating scales. Correlation coefficients for the BDAE profile of speech characteristics, in which three raters evaluated 99 subjects, ranged from 0.78 for word finding to 0.90 for phrase length, articulatory agility, and grammatical form (Goodglass, Barresi, & Kaplan, 1983). Although an actual ICC was not calculated, their correlation of the most disparate raters on each scale likely represents good-to-excellent reliability. Inter-rater reliability was also examined for the WAB fluency scale, where 8 judges evaluated 10 patients of varying types and severity (Kertesz, 1982). Average intercorrelations were reported to be 0.98. While this report is remarkably high, the unidimensional nature of the scale likely limited the potential variability in scoring. Inter-rater reliability of a novel, seven-point rating scale created by Wagenaar, Snow, & Prins (1975) was additionally established using the Kendall coefficient of concordance. Coefficients were found to be within the excellent range (α ranged from 0.864 to 0.941), though only four of the thirty variables (communicative capacity, syntactic complexity, melody, articulation) were examined.

Finally, the reliability of the auditory-perceptual approach to motor speech assessment was recently established, where 20 raters evaluated 47 patients of varying dysarthria types on 38 perceptual features (Bunton et al., 2007). Differences between speakers accounted for 36% to 62% of the variance, corresponding to a partial *R* from 0.60 to 0.79. This is comparable to an ICC and is likely representative of reliability in the good-to-excellent range. Interestingly, no significant difference between inexperienced and experienced raters was found by Bunton and colleagues, suggesting that the significant differences observed between rater groups in our study may be the result of inherent differences in rating linguistic versus speech features.

## Validity

Most of the APROCSA features showed good concurrent validity relative to quantitative measures of connected speech, motor speech diagnoses, or WAB subscores. Some of the weaker correlations likely reflect the ambiguity of some of the features in the APROCSA. For instance, the absent correlation between *phonemic paraphasias* and phonological errors (phw) may partially be the result of the relatively low reliability of *phonemic paraphasias* and the difficulty distinguishing phonemic paraphasias from errors resulting from apraxia of speech, as previously discussed.

Other weaker correlations may be due to differences in specificity between the APROCSA features and the AphasiaBank measures. For instance, in the APROCSA, apraxia of speech was rated directly through its own diagnostic feature and indirectly through other features (e.g., *reduced speech rate, halting and effortful*). In contrast, apraxia of speech was not tested directly through AphasiaBank protocol. Instead, examiners from the AphasiaBank documented the presence or absence of apraxia of speech on a binary scale when collecting demographic information.

A mismatch in construct criteria between APROCSA features and CHAT transcription variables may have also played a role. The absence of a correlation between *semantic paraphasias* and semantic errors (phw) was likely the result of differences in how semantic paraphasias were categorized in CHAT. While our manual stipulates that the rater must make a judgment as to whether an error is phonological or semantic in nature, CHAT coding for semantic errors with an unknown target, [*s:uk], makes no distinction regarding the nature of the error. This particular code is defined as an error that results in a real word with an unknown target (MacWhinney, 2000). Given that both semantic and phonemic paraphasias may result in a real word with an unknown target, either may be labelled as such using the CHAT coding scheme. As a result, phonemic paraphasias may be inadvertently labeled with this code, thereby confounding the correlation analysis. Similarly, *conduite d'approche* likely failed to correlate with the CHAT transcription measures retraced sequences (phw) and false starts (phw), as neither was a direct measure of the construct. While retracing and false starts may be the result of conduite d'approche, they often occur in the absence of conduite d'approche as well.

## Patterns

As expected, many features patterned together with strong correlations within identified APROCSA categories. Unanticipated anticorrelations (i.e., negative correlations) were also observed among the features. For instance, *omission of bound morphemes* and *omission of* *function words* were both anticorrelated with *semantic paraphasias*. It is unclear whether these findings are spurious or reflective of patterns that warrant further investigation.

Four readily interpretable underlying factors were shown to account for much of the variance across the 27 connected speech features. One factor loaded on phenomena associated with fluent aphasia (Paraphasia), while the other three—Logopenia, Agrammatism and Motor speech—reflected a parcellation of dimensions of non-fluency. Much previous research has shown that aspects of non-fluency can dissociate (Goodglass, Quadfasel, & Timberlake, 1964; Benson, 1967; Wilson et al., 2010; Thompson et al., 2012), but our data-driven identification of precisely three specific dimensions including a Logopenia dimension is intriguing. Individual patients presented with varying mixtures of the three non-fluent factors, as well as the fluent factor (Paraphasia), which was not simply the opposite of non-fluency, but could occur in conjunction with the non-fluent dimensions. Factor loadings by patient were varied within a given WAB subtype, with multiple factor loadings often observed within a single patient. This variation within a given WAB subtype has been previously documented, showing dissociation of a given WAB subtype and grouping of multiple WAB subtypes within a single cluster pattern (Kertesz & Phipps, 1977).

It is difficult to ascertain the extent to which the observed factors were determined by our cohort of patients with chronic post-stroke aphasia. In a cohort of patients with acute post-stroke aphasia where motor and linguistic deficits commonly co-occur, we may expect to see dissociation of the Motor speech factor, with one representing apraxia of speech and the other dysarthria. A cohort of patients with primary progressive aphasia may result in parcellation of the Paraphasia factor into two separate factors, one characterized by semantic and phonemic paraphasias, and the other representative of nonspecific and empty speech.

### Future directions

The results observed in this study warrant further investigation into the APROCSA as a clinical and research tool. As mentioned above, one potential area of research is the administration of the APROCSA with different cohorts, such as acute post-stroke aphasia or primary progressive aphasia, to examine whether the factors observed in this study were cohort-specific or generalized behaviors in aphasia. Another possible avenue is to determine the reliability and validity of rating the four factors derived from the APROCSA directly, as opposed to rating the 27 connected speech features. Finally, quantifying correlations between APROCSA-derived variables and factors and neuroimaging data is an important next step in determining whether the observed behaviors follow the neuroanatomy and neurophysiology of patients with aphasia of differing etiologies.

## References

Albert, M. L., Obler, L. K., Goodglass, H., Helm, N. A., Rubens, A., & Alexander, M. P. (1981). *Clinical aspects of dysphasia*. Wien/New York: Springer Verlag.

Benson, D. F. (1967). Fluency in aphasia: Correlation with radioactive scan localization. *Cortex, 3*, 373–394.

Boles, L. & Bombard, T. (1998). Conversational discourse analysis: appropriate and useful sample sizes. *Aphasiology, 12*, 547–560.

Bunton, K., Kent, R. D., Duffy, J. R., Rosenbek, J. C., & Kent, J. F. (2007). Listener agreement for auditory-perceptual ratings of dysarthria. *Journal of Speech, Language, and Hearing Research, 50,* 1481–1495.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6,* 284–290.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research*, *12,* 246–269.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech, Language, and Hearing Research*, *12,* 462–496.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders.* Philadelphia: Saunders.

Duffy, J. R. (2013). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management* (3rd ed.). St. Louis: Elsevier/Mosby.

Goodglass, H., Quadfasel, F. A., Timberlake, W. H. (1964). Phrase length and the type of severity of aphasia. *Cortex, 1,* 133–153.

Goodglass, H., Barresi, B., & Kaplan, E. (1983). *The Boston Diagnostic Aphasia Examination*

(2nd ed.). Lippincott Williams & Wilkins.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *The Boston Diagnostic Aphasia Examination (BDAE)* (3rd ed.). Baltimore: Lippincott Williams & Wilkins.

Kertesz, A. (1979). *Aphasia and associated disorders: Taxonomy, localization, and recovery*. New York: Grune & Stratton.

Kertesz, A. (1982). *Western Aphasia Battery.* New York: Grune & Stratton.

Kertesz, A., & Phipps, J. B. (1977). Numerical taxonomy of aphasia. *Brain and Language, 4*(1), 1–10.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Lawrence Erlbaum.

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

McCarron, A., Chavez, A., Babiak M., Berger, M. S., Chang, E. F., & Wilson, S. M. (2017). Connected speech in transient aphasias after left hemisphere resective surgery. *Aphasiology,* Forthcoming.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, *1*(1), 30–46.

Prins, R., & Bastiaanse, R. (2004). Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, *18,* 1075–1091.

Prins, R. S., Snow, C. E., & Wagenaar, E. (1978). Recovery from aphasia: Spontaneous speech versus language comprehension. *Brain and Language*, *6*(2), 192–211.

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,* 440–479.

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.

Thompson, C. K., Cho, S., Hsu, C.-J., Wieneke, C., Rademaker, A., Weitner, B. B., … Weintraub, S. (2012). Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology, 26,* 20–43.

Wagenaar, E., Snow, C., & Prins, R. (1975). Spontaneous speech of aphasic patients: A psycholinguistic analysis. *Brain and language*, *2,* 281–303.

Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., … Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain, 133,* 2069–2088.

Yagata, S. A., Yen, M., McCarron, A., Bautista, A., Lamair-Orosco, G., & Wilson, S. M. (2017). Rapid recovery from aphasia after infarction of Wernicke's area. *Aphasiology,* Forthcoming.

**Table 1.** The 27 features of the APROCSA

| Connected speech features | Definition |
|---|---|
| *Lexical retrieval* | |
| Anomia | Overall impression of word-finding difficulties |
| Abandoned utterances | Utterances are left incomplete |
| Empty speech | Speech that conveys little or no meaning |
| *Selection of words and sounds* | |
| Semantic paraphasias | Substitution of a content word for a related or unrelated content word |
| Phonemic paraphasias | Substitution, insertion, deletion, or transposition of one or two clearly articulated phonemes |
| Neologisms | Word forms that are not real English words |
| Jargon | Fluent, prosodically correct but meaningless speech |
| Perseverations | Repetition of a previously used word or utterance |
| Stereotypies | Commonly used words or phrases produced with relative ease and fluency |
| *Grammatical construction* | |
| Short and simplified utterances | Speech is reduced in length or complexity |
| Omission of bound morphemes | Inflectional or derivational morphemes are not used where they should be |
| Omission of function words | Function words are not used where they should be |
| Paragrammatism | Inappropriate juxtaposition or misuse of words |
| *Rate and timing* | |
| Pauses between utterances | Pauses between the speaker's utterances or responses to the examiner's questions |
| Pauses within utterances | Filled (*um*, *uh*) or silent pauses within an utterance |
| Halting and effortful | Prosody or melodic line is disrupted or unnatural |
| Reduced speech rate | Rate in typical sequences is slower than expected |
| *Self-correction* | |
| False starts | Partial words are abandoned after a few phonemes |
| Retracing | Sequences of one or more complete words, which are made redundant by subsequent revisions |
| Conduite d'approche | Successive attempts at an apparent target form |
| *Clarity* | |
| Target unclear | It is not clear what phonemes the speaker is |
| Meaning unclear | The context of the speaker's utterances is unclear |
| Off-topic | The speaker's utterances are clear but out of context |
| *Diagnostic* | |
| Expressive aphasia | Language production is disrupted |
| Apraxia of speech | Speech contains distortions, substitutions, or omissions that tend to increase with length or complexity of the word or phrase |
| Dysarthria | Speech is difficult to understand and characterized as *slurred*, *choppy*, or *mumbled* |
| Overall communication impairment | Extent to which the speaker exhibits difficulty conveying their message |

**Table 2.** The 5-point rating scale used in the APROCSA

| Score | Severity | Description |
|---|---|---|
| 0 | Not present | Not present or within the range of healthy older speakers |
| 1 | Mild | Detectable but infrequent |
| 2 | Moderate | Frequently evident but not pervasive |
| 3 | Marked | Moderately severe, pervasive |
| 4 | Severe | Nearly always evident |

The scale is based on Strand, Duffy, Clark, & Josephs (2014).

**Table 3.** Patient characteristics

| Patient | Age (years) | Sex | Race | Education (years) | Duration post-onset (months) | WAB-AQ (out of 100) | BNT short form (out of 15) | Aphasia subtype | Apraxia | Dysarthria |
|---|---|---|---|---|---|---|---|---|---|---|
| fridriksson05a | 58.3 | F | WH | 12 | 149 | 92.7 | 13 | Anomic | Y | N |
| TAP18a | 53.7 | F | WH | 16 | 23 | 90.3 | 12 | Anomic | Y | N |
| whiteside06a | 62 | M | WH | 12 | 91 | 88.8 | 8 | Anomic | Y | N |
| adler01a | 58.8 | M | WH | 13 | 16 | 86.8 | 12 | Anomic | Y | Y |
| kurland07a | 70.6 | F | WH | 16 | 13 | 83 | 11 | Anomic | N | N |
| kurland28a | 62.5 | M | WH | 16 | 6 | 78.7 | 4 | Anomic | N | N |
| scale30a | 48.9 | M | WH | 18 | 46 | 68.5 | 7 | Anomic | N | N |
| ACWT09a | 56.2 | F | WH | 13 | 94 | 80.1 | 11 | Conduction | Y | N |
| wright203a | 66.4 | M | WH | 18 | 80 | 76.3 | 11 | Conduction | N | N |
| williamson04a | 60.9 | M | WH | 14 | 296 | 70.6 | 2 | Conduction | Y | N |
| kurland20a | 50.1 | F | AA | 12 | 6 | 67 | 6 | Conduction | N | N |
| TCU07a | 49.2 | F | WH | 16 | 15 | 52 | 1 | Conduction | Y | N |
| williamson16a | 63.5 | F | WH | 16 | 58 | 66.4 | 2 | Trans Sensory | N | N |
| ACWT02a | 53.1 | F | WH | 14 | 39 | 74.6 | 8 | Trans Motor | Y | N |
| elman12a | 57.4 | M | WH | 20 | 54 | 74.4 | 10 | Wernicke | N | N |
| elman14a | 76.3 | F | AA | 17 | 55 | 65.7 | 9 | Wernicke | N | N |
| thompson05a | 63.9 | F | WH | 16 | 155 | 58.5 | 14 | Wernicke | - | - |
| kurland18a | 74.3 | M | AA | 16 | 9 | 44 | 2 | Wernicke | N | N |
| scale33a | 57.3 | F | WH | - | 104 | 71.1 | 7 | Broca | N | N |
| TCU08a | 57.2 | M | AA | 14 | 95 | 63.9 | 4 | Broca | Y | N |
| TAP11a | 62.7 | F | WH | 14 | 44 | 58.1 | 1 | Broca | Y | N |
| BU08a | 64.6 | M | WH | 12 | 110 | 39.7 | 1 | Broca | N | N |
| TAP09a | 71 | M | WH | 16 | 36 | 20.5 | 1 | Global | Y | N |
| scale09a | 66.2 | M | WH | 12 | 240 | 20.3 | 2 | Global | Y | Y |

The - symbol indicates no information was provided.

**Table 4.** Student rater characteristics

| Student rater characteristics | |
|---|---|
| Age | 22 – 33 years (mean: 25.5 ± 3.3) |
| Sex | 11 female, 1 male |
| First language | 10 English, 1 Shanghainese and English, 1 Korean |
| Highest degree earned | 10 Bachelors, 2 Masters |
| Clinical experience in adult language | 25 – 200 hours (mean 86 ± 50 hours) |
| Clinical settings in adult language | University aphasia clinic (all), acute care (5), inpatient rehabilitation (5), private clinic (2) |
| Research experience | 0 – 4220 hours (mean 1099 ± 1211 hours) |
| Transcription experience | 0 – 1920 hours (mean 376 ± 677 hours) |
| Auditory-perceptual experience | 0 – 50 hours (mean 12.5 ± 16.0 hours) |
| Confidence in aphasia | 4 – 5 on a 5-point Likert scale (mean 4.4 ± 0.5) |
| Confidence in motor speech disorders | 2 – 4 on a 5-point Likert scale (mean 3.2 ± 0.7) |
| Graduate course in language disorders | All completed |
| Graduate course in motor speech disorders | 1 completed; 11 in progress |

**Table 5.** Quantitative linguistic measures derived from CHAT and CLAN

| Quantitative linguistic measure | Description |
|---|---|
| Anomia (phw) | Post codes +… +..? for abandoned utterances, [+es] for empty speech, (.) (..) for pausing, and [&ah] [&eh] [&ew] [&hm] [&mm] [&uh] [&uhm] [&um] for filled pauses were summed using FREQ and the proportion per hundred words was taken. |
| Abandoned utterances (phw) | Post codes +… +..? were summed using FREQ and the proportion per hundred words was taken. |
| Empty speech (phw) | Post code [+es] was summed using FREQ and the proportion per hundred words was taken. |
| Semantic errors (phw) | Word-level error codes [*s:r] [*s:ur] [*s:uk] [*s:per] were summed using FREQ and the proportion per hundred words was taken. |
| Phonological errors (phw) | Word-level error codes [*p:w] [*p:m] [*p:n] were summed using FREQ and the proportion per hundred words was taken. |
| Neologisms (phw) | Word-level error codes [*n:k] [*n:uk] were summed using FREQ and the proportion per hundred words was taken. |
| Jargon (phw) | Word-level error codes [*s] for semantic errors, [*p] for phonological errors, and [*n] for neologistic errors were with post code [+jar] using FREQ. The proportion per hundred words was then taken. |
| MLU (morphemes) | MLU in morphemes was calculated using EVAL. Revisions, fillers, and unintelligible utterances were excluded. |
| Bound morphemes (proportion) | %mor line codes for bound morphemes (plurals, 3S, 1S/3S, PAST, PASTP, PRESP) and free morphemes (nouns, verbs, auxiliaries, prepositions, adjectives, adverbs, conjunctions, determiners/articles, pronouns) were summed using EVAL. The proportion of bound-to-free morphemes was then taken. |
| Closed class words (proportion) | %mor line codes for closed class words (auxiliaries, prepositions, conjunctions, determiners/articles, pronouns) and open class words (nouns, verbs, adjectives, adverbs) were summed using EVAL. The proportion of closed-to-open class words was then taken. |
| Pronouns (proportion) | %mor line codes for nouns and pronouns were summed using EVAL. The proportion of pronouns-to-nouns was then taken. |
| Agrammatic utterances (phw) | Post code [+gram] was summed using FREQ and the proportion per hundred words was taken |
| Pauses (phw) | Post codes (.) (..) (...) [&ah] [&eh] [&ew] [&hm] [&mm] [&uh] [&uhm] [&um] were summed using FREQ and the proportion per hundred words was taken. |
| Words per minute | Words per minute was calculated using EVAL and were based on time-stamped codes embedded in the transcript file. |
| Retraced sequences (phw) | Post codes [/] [//] were summed using EVAL and the proportion per hundred words was taken. |
| False starts (phw) | Post code [&] was summed using FREQ and the proportion per hundred words was taken. [&] codes for gestures and filled pauses were not included in the calculation. |
| Unintelligible sequences (phw) | Word-level error code xxx was summed using FREQ and the proportion per hundred words was taken. |

## Figure captions

**Figure 1.** Distribution and reliability of the 27 connected speech variables. Each row shows one connected speech variable. The first column shows the distribution of the 24 patients' scores, where each patient's score is the mean of the three expert ratings. Boxes: interquartile ranges; Whiskers: ranges excluding outliers; circles: outliers; red lines: medians; blue asterisks: means. The second column shows the intraclass correlation coefficient (ICC), type A,k, for the three experts. This is the expected correlation between scores averaged across the three experts, and scores averaged across three different hypothetical experts from the same population of experts. Error bars indicate 95% confidence intervals. The third column shows the ICC, type A,1, for the three experts. This is the expected correlation between individual experts from the population of experts. The fourth column shows the distribution of the 24 patients' scores, where each patient's score is the mean of four student ratings (only 4 of the 12 students rated each patient). Red lines: medians; blue asterisks: means; black circles: outliers. The fifth column shows the ICC, type 1,k, for the students. This is the expected correlation between scores averaged across four students, and scores averaged across a different set of four students, with all students drawn at random from the population of students. The sixth column shows the ICC, type 1, for the students. This is the expected correlation between individual students from the population of students. The ICCs are color-coded poor (<0.40), fair ($0.40 \leq r \leq 0.60$), good ($0.60 \leq r \leq 0.75$) or excellent ($r \geq 0.75$), following Cicchetti (1994).

**Figure 2.** Concurrent validity of the 27 connected speech features. Pearson correlation coefficients are indicated by depth of color, and Pearson *r* values are shown for correlations with uncorrected $p < 0.05$. The *y* axis shows the 27 connected speech features. The *x* axis shows: (1)

25 quantitative measures derived from the transcription and coding of the speech samples in AphasiaBank; (2) two binary motor speech diagnoses reported in AphasiaBank; and (3) the five subscores and the Aphasia Quotient from the Western Aphasia Battery (WAB). APROCSA connected speech features are all defined such that high scores are indicative of impairment. The other measures differ in terms of their directionality. In general, the blue color scale is used to encode correlations of scores indicating impairment with scores indicating impairment, whereas the red color scale is used to encode correlations of scores indicating impairment with scores indicating sparing. Exception to this are three AphasiaBank quantitative measures—bound morphemes (proportion), closed class words (proportion), and pronouns (proportion) —since these measures can be perturbed in either direction in aphasia. The “agrammatic” perturbations of these scores were arbitrarily defined as the direction of impairment. Yellow boxes indicate AphasiaBank measures that were considered *a priori* to be measuring the same or similar phenomena to each connected speech feature.

**Figure 3.** Patterning of connected speech variables: correlation matrix. Each variable is shown on both the x and y axes, so the matrix is symmetric around the diagonal. Positive correlations are indicated in blue and negative correlations in red. Pearson r values are shown for correlations with uncorrected $p < 0.05$.

**Figure 4.** Patterning of connected speech features: factor analysis. Only 23 of the 27 features were used, since there were only 24 patients. A four-factor rotated model provided the most explanatory account of the data. The factors were labeled Paraphasia, Logopenia, Agrammatism and Motor speech. Loadings of each feature on each factor are shown, and accompanied by bars:

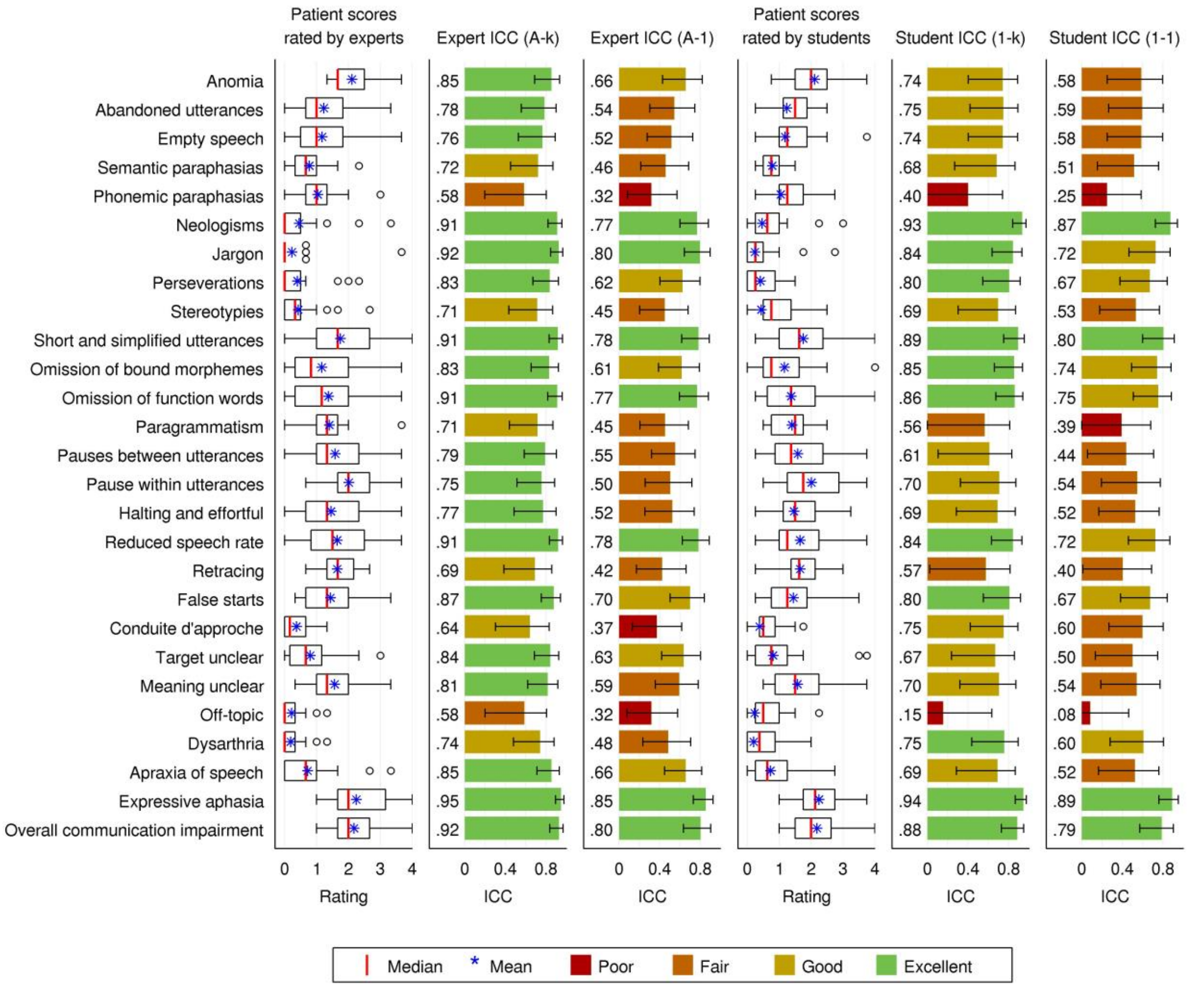
positive in blue and negative in red. Communality indicates the proportion of variance of each feature that is explained by the four factors.

**Figure 5.** Concurrent validity of the four factors. Pearson correlation coefficients are plotted for correlations between each factor and a number of variables from AphasiaBank. See Figure 2 caption for details.

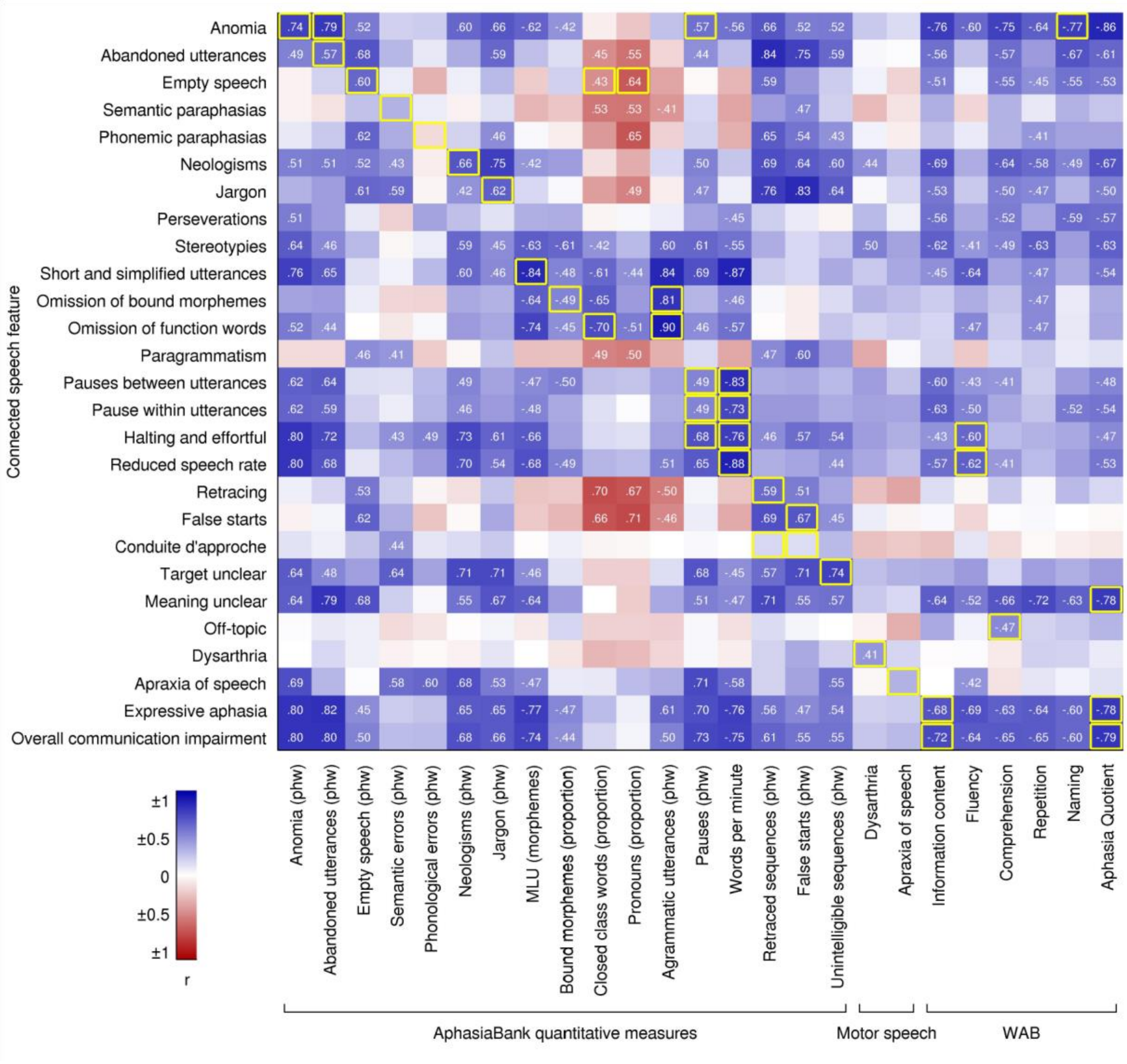
**Figure 6.** Characteristics of individuals with aphasia. For each of the 24 patients, the scores on each of the four factors are shown. Patients are ordered by WAB subtype, with less severe subtypes first, then by descending AQ within subtype. Patients with the same WAB subtype showed different connected speech characteristics.

# Figures

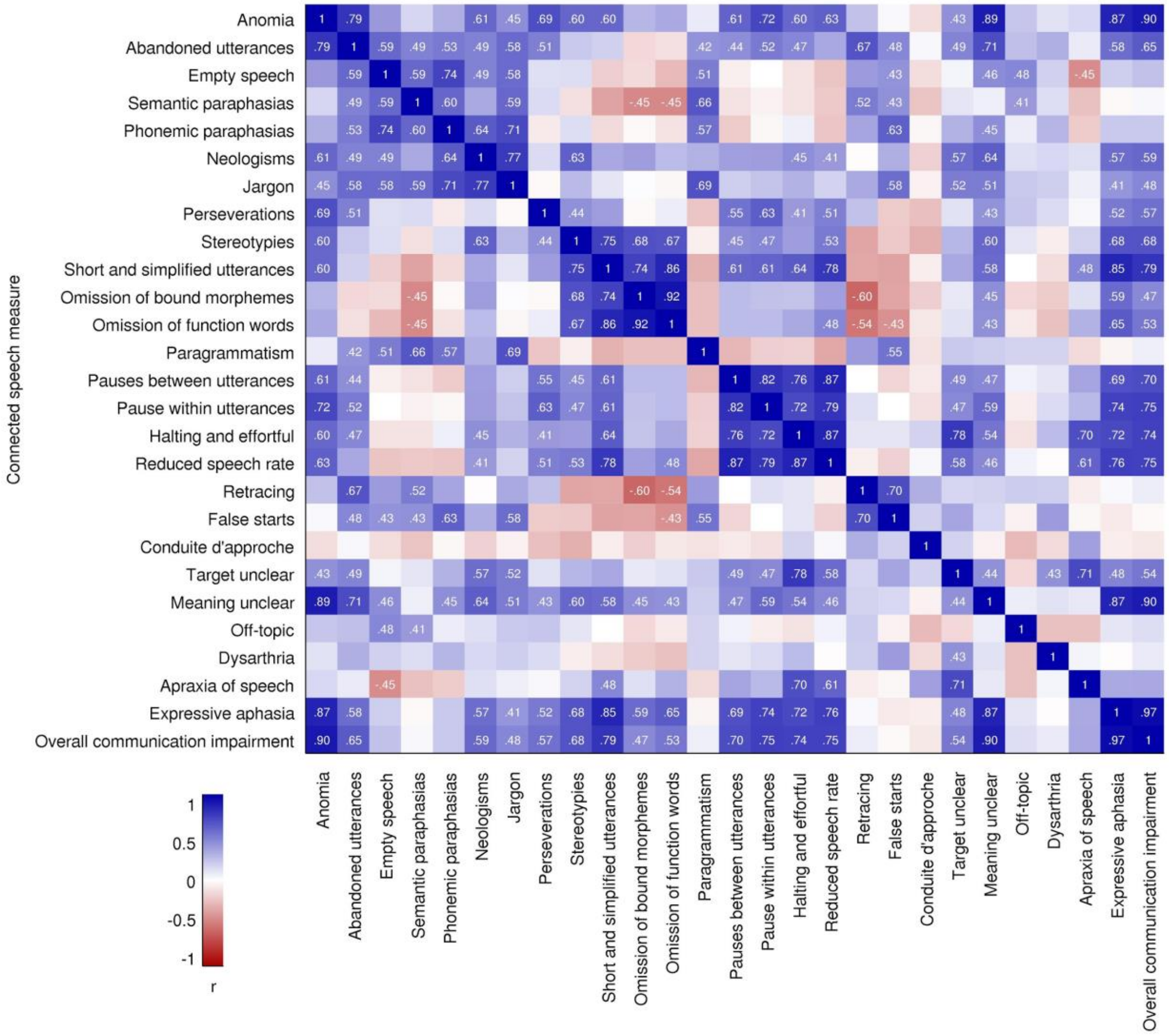
## Figure 1



# Figure 2

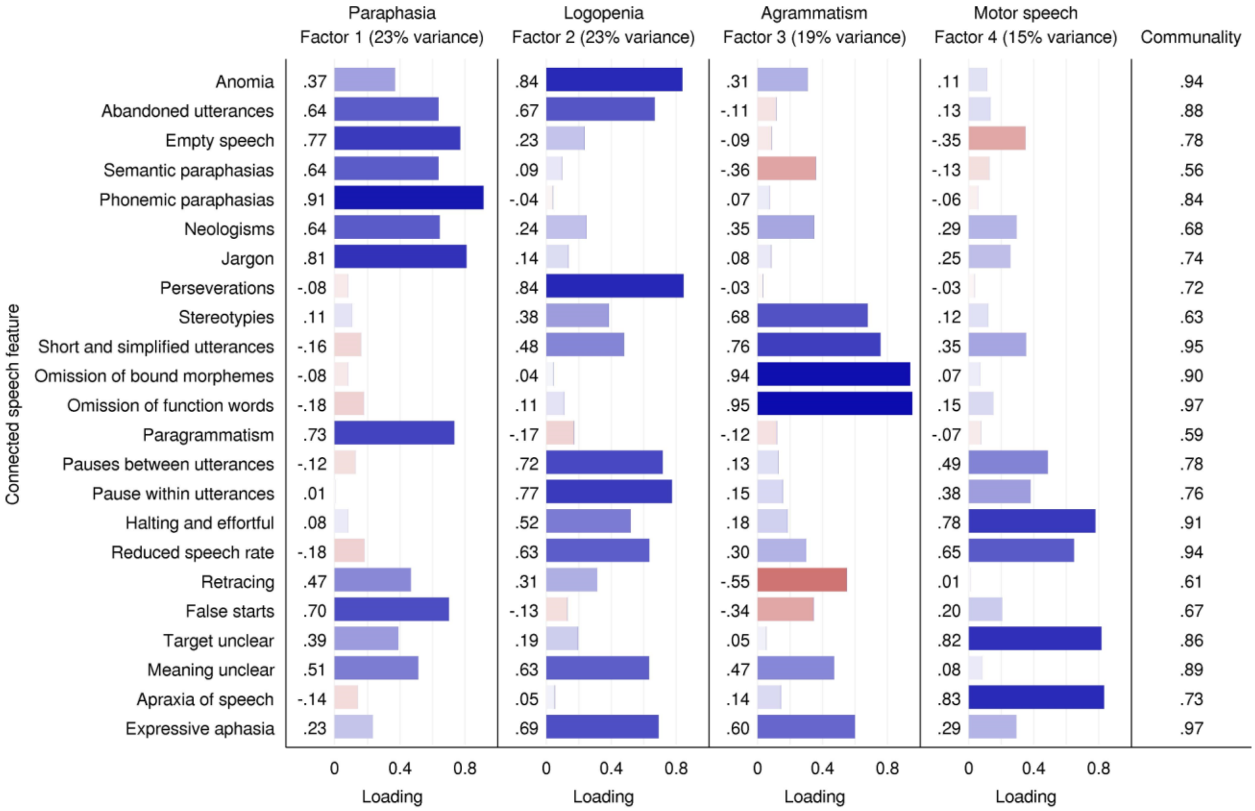

Connected speech feature
Anomia .74 .79 .52 .60 .66 -.62 -.42 .57 -.56 .66 .52 .52 -.76 -.60 -.75 -.64 -.77 -.86
Abandoned utterances .49 .57 .68 .59 .45 .55 .44 .84 .75 .59 -.56 -.57 -.67 -.61
Empty speech .60 .43 .64 .59 -.51 -.55 -.45 -.55 -.53
Semantic paraphasias .53 .53 -.41 .47
Phonemic paraphasias .62 .46 .65 .65 .54 .43 -.41
Neologisms .51 .51 .52 .43 .66 .75 -.42 .50 .69 .64 .60 .44 -.69 -.64 -.58 -.49 -.67
Jargon .61 .59 .42 .62 .49 .47 .76 .83 .64 -.53 -.50 -.47 -.50
Perseverations .51 -.45 -.56 -.52 -.59 -.57
Stereotypies .64 .46 .59 .45 -.63 -.61 -.42 .60 .61 -.55 .50 -.62 -.41 -.49 -.63 -.63
Short and simplified utterances .76 .65 .60 .46 -.84 -.48 -.61 -.44 .84 .69 -.87 -.45 -.64 -.47 -.54
Omission of bound morphemes -.64 -.49 -.65 .81 -.46 -.47
Omission of function words .52 .44 -.74 -.45 -.70 -.51 .90 .46 -.57 -.47 -.47
Paragrammatism .46 .41 .49 .50 .47 .60
Pauses between utterances .62 .64 .49 -.47 -.50 .49 -.83 -.60 -.43 -.41 -.48
Pause within utterances .62 .59 .46 -.48 .49 -.73 -.63 -.50 -.52 -.54
Halting and effortful .80 .72 .43 .49 .73 .61 -.66 .68 -.76 .46 .57 .54 -.43 -.60 -.47
Reduced speech rate .80 .68 .70 .54 -.68 -.49 .51 .65 -.88 .44 -.57 -.62 -.41 -.53
Retracing .53 .70 .67 -.50 .59 .51
False starts .62 .66 .71 -.46 .69 .67 .45
Conduite d'approche .44
Target unclear .64 .48 .64 .71 .71 -.46 .68 -.45 .57 .71 .74
Meaning unclear .64 .79 .68 .55 .67 -.64 .51 -.47 .71 .55 .57 -.64 -.52 -.66 -.72 -.63 -.78
Off-topic -.47
Dysarthria .41
Apraxia of speech .69 .58 .60 .68 .53 -.47 .71 -.58 .55 -.42
Expressive aphasia .80 .82 .45 .65 .65 -.77 -.47 .61 .70 -.76 .56 .47 .54 -.68 -.69 -.63 -.64 -.60 -.78
Overall communication impairment .80 .80 .50 .68 .66 -.74 -.44 .50 .73 -.75 .61 .55 .55 -.72 -.64 -.65 -.65 -.60 -.79
±1
±0.5
0
±0.5
±1
r
Anomia (phw)
Abandoned utterances (phw)
Empty speech (phw)
Semantic errors (phw)
Phonological errors (phw)
Neologisms (phw)
Jargon (phw)
MLU (morphemes)
Bound morphemes (proportion)
Closed class words (proportion)
Pronouns (proportion)
Agrammatic utterances (phw)
Pauses (phw)
Words per minute
Retraced sequences (phw)
False starts (phw)
Unintelligible sequences (phw)
Dysarthria
Apraxia of speech
Information content
Fluency
Comprehension
Repetition
Naming
Aphasia Quotient
AphasiaBank quantitative measures
Motor speech
WAB

## Figure 3



| | Anomia | Abandoned utterances | Empty speech | Semantic paraphasias | Phonemic paraphasias | Neologisms | Jargon | Perseverations | Stereotypies | Short and simplified utterances | Omission of bound morphemes | Omission of function words | Paragrammatism | Pauses between utterances | Pause within utterances | Halting and effortful | Reduced speech rate | Retracing | False starts | Conduite d'approche | Target unclear | Meaning unclear | Off-topic | Dysarthria | Apraxia of speech | Expressive aphasia | Overall communication impairment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Anomia | 1 | .79 | | | | .61 | .45 | .69 | .60 | .60 | | | | .61 | .72 | .60 | .63 | | | | .43 | .89 | | | | .87 | .90 |
| Abandoned utterances | .79 | 1 | .59 | .49 | .53 | .49 | .58 | .51 | | | | | .42 | .44 | .52 | .47 | | .67 | .48 | | .49 | .71 | | | | .58 | .65 |
| Empty speech | | .59 | 1 | .59 | .74 | .49 | .58 | | | | | | .51 | | | | | | .43 | | | .46 | .48 | | -.45 | | |
| Semantic paraphasias | | .49 | .59 | 1 | .60 | | .59 | | | | -.45 | -.45 | .66 | | | | | .52 | .43 | | | | .41 | | | | |
| Phonemic paraphasias | | .53 | .74 | .60 | 1 | .64 | .71 | | | | | | .57 | | | | | | .63 | | | .45 | | | | | |
| Neologisms | .61 | .49 | .49 | | .64 | 1 | .77 | | .63 | | | | | | | .45 | .41 | | | | .57 | .64 | | | | .57 | .59 |
| Jargon | .45 | .58 | .58 | .59 | .71 | .77 | 1 | | | | | | .69 | | | | | | .58 | | .52 | .51 | | | | .41 | .48 |
| Perseverations | .69 | .51 | | | | | | 1 | .44 | | | | | .55 | .63 | .41 | .51 | | | | | .43 | | | | .52 | .57 |
| Stereotypies | .60 | | | | | .63 | | .44 | 1 | .75 | .68 | .67 | | .45 | .47 | | .53 | | | | | .60 | | | | .68 | .68 |
| Short and simplified utterances | .60 | | | | | | | | .75 | 1 | .74 | .86 | | .61 | .61 | .64 | .78 | | | | | .58 | | | .48 | .85 | .79 |
| Omission of bound morphemes | | | | -.45 | | | | | .68 | .74 | 1 | .92 | | | | | | -.60 | | | | .45 | | | | .59 | .47 |
| Omission of function words | | | | -.45 | | | | | .67 | .86 | .92 | 1 | | | | | .48 | -.54 | -.43 | | | .43 | | | | .65 | .53 |
| Paragrammatism | | .42 | .51 | .66 | .57 | | .69 | | | | | | 1 | | | | | | .55 | | | | | | | | |
| Pauses between utterances | .61 | .44 | | | | | | .55 | .45 | .61 | | | | 1 | .82 | .76 | .87 | | | | .49 | .47 | | | | .69 | .70 |
| Pause within utterances | .72 | .52 | | | | | | .63 | .47 | .61 | | | | .82 | 1 | .72 | .79 | | | | .47 | .59 | | | | .74 | .75 |
| Halting and effortful | .60 | .47 | | | | .45 | | .41 | | .64 | | | | .76 | .72 | 1 | .87 | | | | .78 | .54 | | | .70 | .72 | .74 |
| Reduced speech rate | .63 | | | | | .41 | | .51 | .53 | .78 | | .48 | | .87 | .79 | .87 | 1 | | | | .58 | .46 | | | .61 | .76 | .75 |
| Retracing | | .67 | | .52 | | | | | | | -.60 | -.54 | | | | | | 1 | .70 | | | | | | | | |
| False starts | | .48 | .43 | .43 | .63 | | .58 | | | | | -.43 | .55 | | | | | .70 | 1 | | | | | | | | |
| Conduite d'approche | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | |
| Target unclear | .43 | .49 | | | | .57 | .52 | | | | | | | .49 | .47 | .78 | .58 | | | | 1 | .44 | | .43 | .71 | .48 | .54 |
| Meaning unclear | .89 | .71 | .46 | | .45 | .64 | .51 | .43 | .60 | .58 | .45 | .43 | | .47 | .59 | .54 | .46 | | | | .44 | 1 | | | | .87 | .90 |
| Off-topic | | | .48 | .41 | | | | | | | | | | | | | | | | | | | 1 | | | | |
| Dysarthria | | | | | | | | | | | | | | | | | | | | | .43 | | | 1 | | | |
| Apraxia of speech | | | -.45 | | | | | | | .48 | | | | | | .70 | .61 | | | | .71 | | | | 1 | | |
| Expressive aphasia | .87 | .58 | | | | .57 | .41 | .52 | .68 | .85 | .59 | .65 | | .69 | .74 | .72 | .76 | | | | .48 | .87 | | | | 1 | .97 |
| Overall communication impairment | .90 | .65 | | | | .59 | .48 | .57 | .68 | .79 | .47 | .53 | | .70 | .75 | .74 | .75 | | | | .54 | .90 | | | | .97 | 1 |

## Figure 4


Paraphasia
Factor 1 (23% variance)
Logopenia
Factor 2 (23% variance)
Agrammatism
Factor 3 (19% variance)
Motor speech
Factor 4 (15% variance)
Communality
Connected speech feature
Anomia .37 .84 .31 .11 .94
Abandoned utterances .64 .67 -.11 .13 .88
Empty speech .77 .23 -.09 -.35 .78
Semantic paraphasias .64 .09 -.36 -.13 .56
Phonemic paraphasias .91 -.04 .07 -.06 .84
Neologisms .64 .24 .35 .29 .68
Jargon .81 .14 .08 .25 .74
Perseverations -.08 .84 -.03 -.03 .72
Stereotypies .11 .38 .68 .12 .63
Short and simplified utterances -.16 .48 .76 .35 .95
Omission of bound morphemes -.08 .04 .94 .07 .90
Omission of function words -.18 .11 .95 .15 .97
Paragrammatism .73 -.17 -.12 -.07 .59
Pauses between utterances -.12 .72 .13 .49 .78
Pause within utterances .01 .77 .15 .38 .76
Halting and effortful .08 .52 .18 .78 .91
Reduced speech rate -.18 .63 .30 .65 .94
Retracing .47 .31 -.55 .01 .61
False starts .70 -.13 -.34 .20 .67
Target unclear .39 .19 .05 .82 .86
Meaning unclear .51 .63 .47 .08 .89
Apraxia of speech -.14 .05 .14 .83 .73
Expressive aphasia .23 .69 .60 .29 .97
0
0.4
0.8
Loading

# Figure 5

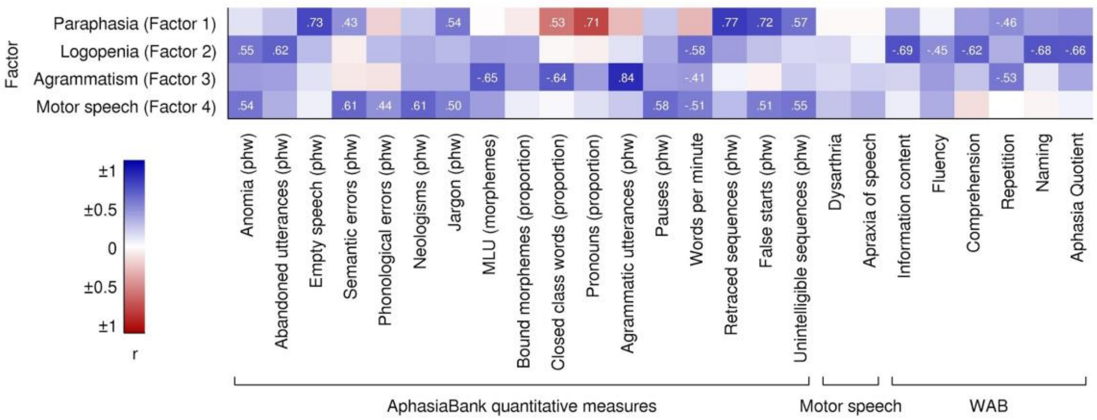
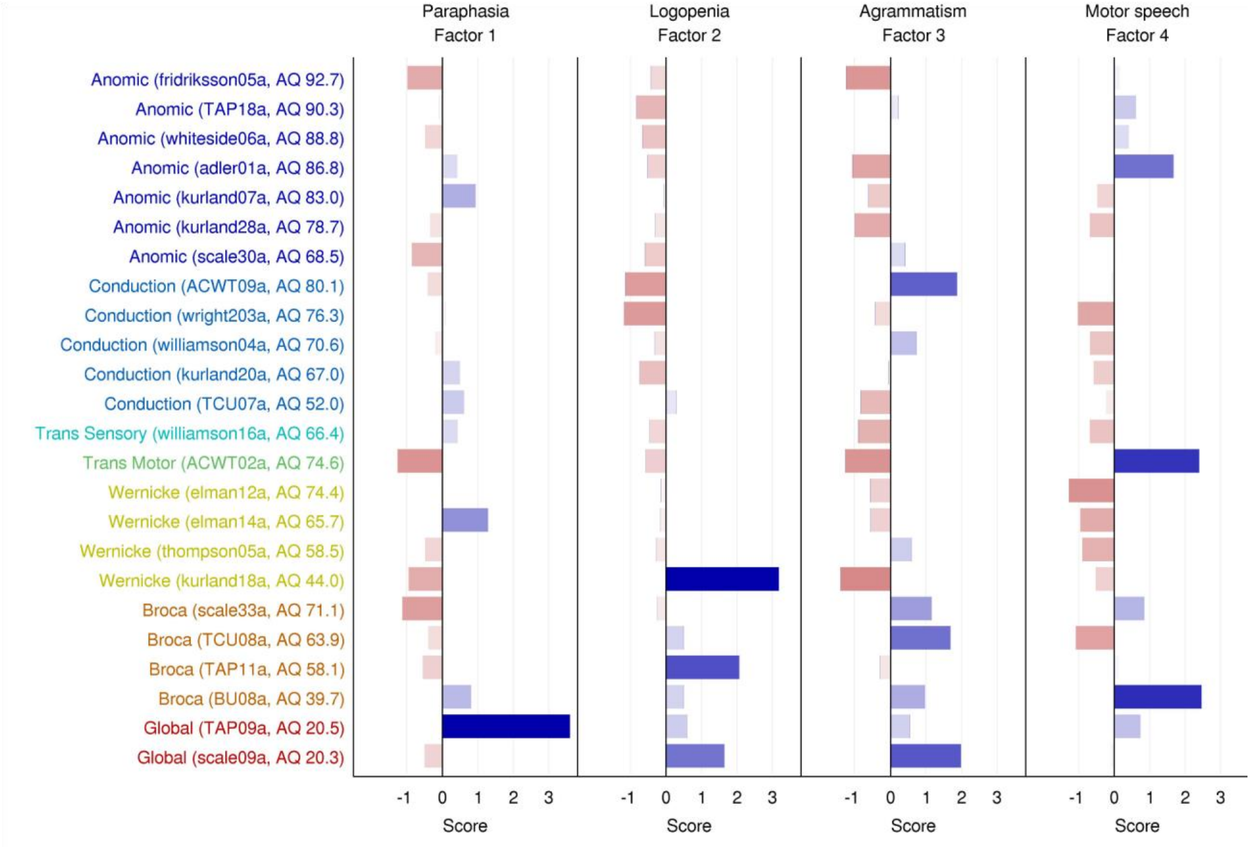

Factor
Paraphasia (Factor 1)
Logopenia (Factor 2)
Agrammatism (Factor 3)
Motor speech (Factor 4)
.73
.43
.54
.53
.71
.77
.72
.57
-.46
.55
.62
-.58
-.69
-.45
-.62
-.68
-.66
-.65
-.64
.84
-.41
-.53
.54
.61
.44
.61
.50
.58
-.51
.51
.55
Anomia (phw)
Abandoned utterances (phw)
Empty speech (phw)
Semantic errors (phw)
Phonological errors (phw)
Neologisms (phw)
Jargon (phw)
MLU (morphemes)
Bound morphemes (proportion)
Closed class words (proportion)
Pronouns (proportion)
Agrammatic utterances (phw)
Pauses (phw)
Words per minute
Retraced sequences (phw)
False starts (phw)
Unintelligible sequences (phw)
Dysarthria
Apraxia of speech
Information content
Fluency
Comprehension
Repetition
Naming
Aphasia Quotient
±1
±0.5
0
±0.5
±1
r
AphasiaBank quantitative measures
Motor speech
WAB

**Figure 6**


Paraphasia
Factor 1
Logopenia
Factor 2
Agrammatism
Factor 3
Motor speech
Factor 4
Anomic (fridriksson05a, AQ 92.7)
Anomic (TAP18a, AQ 90.3)
Anomic (whiteside06a, AQ 88.8)
Anomic (adler01a, AQ 86.8)
Anomic (kurland07a, AQ 83.0)
Anomic (kurland28a, AQ 78.7)
Anomic (scale30a, AQ 68.5)
Conduction (ACWT09a, AQ 80.1)
Conduction (wright203a, AQ 76.3)
Conduction (williamson04a, AQ 70.6)
Conduction (kurland20a, AQ 67.0)
Conduction (TCU07a, AQ 52.0)
Trans Sensory (williamson16a, AQ 66.4)
Trans Motor (ACWT02a, AQ 74.6)
Wernicke (elman12a, AQ 74.4)
Wernicke (elman14a, AQ 65.7)
Wernicke (thompson05a, AQ 58.5)
Wernicke (kurland18a, AQ 44.0)
Broca (scale33a, AQ 71.1)
Broca (TCU08a, AQ 63.9)
Broca (TAP11a, AQ 58.1)
Broca (BU08a, AQ 39.7)
Global (TAP09a, AQ 20.5)
Global (scale09a, AQ 20.3)
-1
0
1
2
3
Score
-1
0
1
2
3
Score
-1
0
1
2
3
Score
-1
0
1
2
3
Score

**Appendix 1.** The APROCSA rating form

Rate connected speech using the following scale:

**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

| **Connected Speech Features** | **0** | **1** | **2** | **3** | **4** |
|---|---|---|---|---|---|
| *Lexical retrieval* | | | | | |
| Anomia | not present | mild | moderate | marked | severe |
| Abandoned utterances | not present | mild | moderate | marked | severe |
| Empty speech | not present | mild | moderate | marked | severe |
| *Selection of words and sounds* | | | | | |
| 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 |
| *Grammatical construction* | | | | | |
| 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 |
| *Rate and timing* | 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 |
| *Self-correction* | | | | | |
| Retracing | not present | mild | moderate | marked | severe |
| False starts | not present | mild | moderate | marked | severe |
| Conduite d’approche | not present | mild | moderate | marked | severe |
| *Clarity* | | | | | |
| Target unclear | not present | mild | moderate | marked | severe |
| Meaning unclear | not present | mild | moderate | marked | severe |
| Off-topic | not present | mild | moderate | marked | severe |
| *Diagnostic category* | | | | | |
| 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 |

## **Appendix 2.** The APROCSA manual

The Auditory-Perceptual Rating of Connected Speech in Aphasia (APROCSA) is a multidimensional rating scheme designed to comprehensively assess the presence and severity of common characteristics of connected speech in aphasia. The features are representative of speech-language impairments that manifest in aphasia of all etiologies or typologies. Collectively, the connected speech features are representative of all language domains (i.e., phonology, morphology, syntax, semantics). A few features additionally represent the speech subsystems (i.e., respiration, phonation, resonance, articulation).

The APROCSA consists of 27 connected speech features that are each scored using a five-point scale, similar to the rating systems developed for dysarthria and apraxia (Darley, Aronson, & Brown, 1969; Strand et al., 2014). Terms and definitions for the 5-point scale are as follows:

**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 frequent but not pervasive
**Marked (3)** = moderately severe impairment or pervasive
**Severe (4)** = severe impairment or nearly always evident

More specific guidelines for certain connected speech features are described below.

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. It is not uncommon for healthy speakers to produce occasional paragrammatic utterances or to abandon utterances. Consequently, if an individual with aphasia exhibits a dimension that would be considered within the expected bounds for a healthy non-elderly person, rate the dimension with a score of not present (0).

Furthermore, the connected speech samples collected for this experiment may not represent the full spectrum of aphasia severity, particularly those with more severe aphasia. However, the APROCSA is designed to capture aphasia severity for all individuals with aphasia. Consequently, always consider the 5-point scale within the context of aphasia severity overall, not simply those with aphasia in these selected speech samples.

In some forms of aphasia, patients will attempt to repair their errors. Errors should still be counted as contributing to the relevant dimension even if they are successfully repaired. Repairs will generally contribute to one or more of the *retracing*, *false starts*, or *conduite d'approche* features.

Try to be as objective as possible in rating each dimension. Regardless of whether you think the dimension directly reflects an underlying impairment, or is secondary to some other linguistic, cognitive, or motor process; simply rate what is present in the sample.

Also, consider the features within the context of an utterance. While determining an utterance can be somewhat subjective, it is an important variable that reflects the length and complexity of a person's speech. Consider the following factors: a sentence is an utterance; sentences conjoined with *and* are separate utterances; falling intonation suggests the end of an utterance; and pauses are unreliable markers of utterance boundaries in people with aphasia.

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

**Directions**

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

1. Listen to the sample once. As you listen, rate features as appropriate and take notes on behaviors observed. Do not pause the video recording. Score the protocol online as you would if you were in a clinical setting.

2. Review your scores and notes. Refer to the descriptions of the connected speech features as needed.

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

**Connected Speech Features**

The following is a list of the connected speech features and their corresponding definitions. All of the features are arranged into categories, which are meant to serve as a guideline while rating. Keep in mind that features will often interconnect within and across categories.

| **Features** | **Description and Comments** |
|---|---|
| *Lexical retrieval* | |
| Anomia | Overall impression of word-finding difficulties, which can be instantiated in many different ways: word-finding pauses typically before nouns, and to a lesser extent, verbs; abandoning utterances after failing to retrieve a word; commenting on the inability to retrieve or say words; empty speech; circumlocution. These specific behaviors are scored on their own scales. |
| Abandoned utterances | Utterances are left incomplete. The speaker may move on to another idea, stop speaking, attempt to use another modality (e.g., gesture), give a vague conclusion to the utterance (e.g., shrugs shoulders and say, *you know*), or explicitly comment that they *can't think of the word*, *can't say it*, etc. |
| Empty speech | Speech that conveys little or no meaning due to lack of specificity. Pronouns or general words such as *thing*, *stuff* or *do* are substituted for content words. |
| *Selection of words and sounds* | |
| Semantic paraphasias | Substitution of a content word for a related or unrelated content word (e.g., *dog* for *cat*). Sometimes phonemic paraphasias can result in real words. If the rater believes the paraphasia to be phonemic in origin, score it as such. |
| Phonemic errors | Substitution, insertion, deletion, or transposition of one or two phonemes (e.g., *papple* for *apple*). The target is usually apparent. Phonemic paraphasias involving more than two phonemes should generally be considered neologisms instead. Despite being misordered or incorrect, phonemes should be correctly articulated (i.e., not distorted), unless there is coexisting dysarthria or apraxia of speech. If you believe that there is a coexisting motor speech impairment, try to quantify phonological errors here and motor errors in the *Dysarthria* and *Apraxia of speech* features. |
| Neologisms | Word forms that are not real English words due to substitution, insertion, deletion, or transposition of multiple phonemes. The target may or may not be apparent. |
| Jargon | Mostly fluent and prosodically correct but largely meaningless speech that contains paraphasias, neologisms, or unintelligible strings. Resembles English syntax and inflection. |

Perseverations

Repetition of a previously used word or utterance in a context where it is no longer appropriate.

Stereotypies

Commonly used words or phrases are produced with relative ease and fluency (e.g., *'goddamit!'*). May also be recurring neologisms or non-words.

*Grammatical construction*

Short or simplified utterances

Speech is reduced in length or complexity. A mild rating (1) should reflect utterances that are sometimes shorter than expected based on the context (e.g., simple sentence structures, lack of subordinate clauses). A severe rating (4) should be reserved for single-word utterances. Non-sentence responses (e.g. ***yes***, or *who did you come with? **My wife**.*) should not be considered.

Omission of bound morphemes

Inflectional (e.g.., *work**ed***, *slow**est***) or derivational (***dis**honest*, *drink**able***) morphemes are not used when they should be. 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 only contains single-word utterances that have bound morphemes. A severe rating (4) should be given for speech that is exclusively uninflected single-word utterances.

Omission of function words

Function words (e.g., determiners, prepositions, pronouns, conjunctions, auxiliaries) are not used when they should be. Omission of these elements generally result in ungrammatical utterances (e.g., *I going to the store*). A severe rating (4) should be given for speech that only contains 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*).

*Rate and timing*

Pauses between utterances

Pauses between the question of an examiner and the response of the speaker, as well as pauses between the speaker's utterances. Failure to respond at all, or failure to fluently string together multiple utterances, can be scored here. This dimension may affect scores in other features, such as *Anomia*, *Halting and effortful speech production*, *Reduced speech rate*, etc.

Pauses within utterances

Pauses may be filled (e.g., *um, uh*) or silent. Both prevalence and length of pauses should be taken into account in assessing severity. A small number of pauses, filled or unfilled, should be scored as not present (0).

Halting and effortful

Prosody or melodic line is disrupted and lacks a natural contour. Intonation, rhythm, or stress patterns may be reduced, absent, or inappropriately placed.

Reduced speech rate

The person's speech rate (i.e., speech production with consideration of pauses) in typical sequences of speech is not within the expected bounds of a healthy, older person. Stereotypies should not be considered.

*Self-correction*

Retracing

Sequences of one or more complete words, which are made redundant by subsequent repetitions, amendments, elaborations, or alternative expressions. Retracing may occur at any point within an utterance and can be of varying lengths (i.e., one word to whole phrases). An example is *I, I, I was, I went to the store*.

False starts

Partial words that are abandoned after one or two phonemes have been produced (e.g., *Sh-sh-sh He is 10 years old.*). The speaker may or may not subsequently produce the intended word.

Conduite d'approche

Successive attempts at a clearly apparent target form (e.g., *stun, start, starling, starting* for *startling*). The target may or may not be achieved. The patient is aware of their errors. These instances also contribute to scores for *Retracing*, *Phonemic paraphasias*, or *Neologisms* depending on how close the attempts are to the target.

*Clarity*

Target unclear

It is not clear what phonemes the speaker is attempting to produce, generally because of distortions, apraxia of speech, muttering, mumbling, or in some cases, severe jargon. This dimension captures words or utterances that you would be hard pressed to transcribe in IPA simply because you cannot determine what the target sounds/words might be. In contrast, if you would be able to transcribe their speech sounds/words, but you don't understand their meaning, that would contribute to the *Meaning unclear* dimension.

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. Do not consider the examiner's comments (e.g., paraphrasing or clarification questions) when rating this dimension; rate the clarity of the message based on only the patient's verbal output.

Off-topic

It is clear what the speaker is talking about, but it is not clear how it relates to the context.

*Diagnostic category*

Expressive aphasia

Language production is disrupted; the speaker experiences difficulty expressing oneself. Disruptions may occur across any or all language domains (i.e., phonology, morphology, syntax, semantics). Receptive language should not be considered.

Apraxia of speech

Speech is characterized by distortions, substitutions, or omissions. Errors may or may not be consistent. Errors tend to increase with the length of the word or phrase. Automatic speech (e.g., name, birthday) often contains fewer errors than volitional speech. Groping behaviors or impaired intonation may be present.

Dysarthria

Speech is difficult to understand and can be described 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. The following are intended as guidelines for rating this dimension. 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 everyday topics, but speech-language impairment limits discussion of more complex topics. A marked rating (3) should be used when communication about everyday topics is possible with help from the listener, but the patient shares the burden of communication. A severe rating (4) should be used when all communication is fragmentary, and the listener carries the burden of communication. These guidelines, including some of the specific wording, are based on the BDAE Aphasia Severity Rating Scale.