ASHA American Speech-Language-Hearing Association

AJSLP

Review Article

Connected Speech Fluency in Poststroke and Progressive Aphasia: A Scoping Review of Quantitative Approaches and Features

Claire Cordella,[a] Lauren Di Filippo,[a] Vijaya B. Kolachalama,[b,c] and Swathi Kiran[a]

[a] Department of Speech, Language and Hearing Sciences, Boston University, MA [b] Department of Medicine, Boston University Chobanian & Avedisian School of Medicine, MA [c] Department of Computer Science and Faculty of Computing & Data Sciences, Boston University, MA

ARTICLE INFO

Article History:
Received June 9, 2023
Revision received October 9, 2023
Accepted January 8, 2024

Editor-in-Chief: Katherine C. Hustad
Editor: Jacqueline Laures-Gore

https://doi.org/10.1044/2024_AJSLP-23-00208

ABSTRACT

Purpose: Speech fluency has important diagnostic implications for individuals with poststroke aphasia (PSA) as well as primary progressive aphasia (PPA), and quantitative assessment of connected speech has emerged as a widely used approach across both etiologies. The purpose of this review was to provide a clearer picture on the range, nature, and utility of individual quantitative speech/language measures and methods used to assess connected speech fluency in PSA and PPA, and to compare approaches across etiologies.
Method: We conducted a scoping review of literature published between 2012 and 2022 following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews guidelines. Forty-five studies were included in the review. Literature was charted and summarized by etiology and characteristics of included patient populations and method(s) used for derivation and analysis of speech/language features. For a subset of included articles, we also charted the individual quantitative speech/language features reported and the level of significance of reported results.
Results: Results showed that similar methodological approaches have been used to quantify connected speech fluency in both PSA and PPA. Two hundred nine individual speech-language features were analyzed in total, with low levels of convergence across etiology on specific features but greater agreement on the most salient features. The most useful features for differentiating fluent from nonfluent aphasia in both PSA and PPA were features related to overall speech quantity, speech rate, or grammatical competence.
Conclusions: Data from this review demonstrate the feasibility and utility of quantitative approaches to index connected speech fluency in PSA and PPA. We identified emergent trends toward automated analysis methods and data-driven approaches, which offer promising avenues for clinical translation of quantitative approaches. There is a further need for improved consensus on which subset of individual features might be most clinically useful for assessment and monitoring of fluency.
Supplemental Material: https://doi.org/10.23641/asha.25537237

Among persons with aphasia (PWA), many experience difficulties with verbal expression that render everyday communication effortful, inefficient, and stressful (Cahana-Amitay et al., 2011; Laures-Gore & DeFife, 2013). Verbal expressive impairments are the hallmark characteristic of nonfluent subtypes of aphasia, in contrast to fluent subtypes in which spontaneous connected speech is relatively more preserved though still affected by principal lexical retrieval deficits (National Institute on Deafness and Other Communication Disorders, 2017). Although this fluent/nonfluent distinction is a simplification of the more nuanced clinical symptoms that occur across aphasia subtypes (Poeck, 1989; Tremblay & Dick, 2016; Wilson et al., 2022), it has persisted in part because speech fluency remains highly salient to patients and clinicians. For PWA, nonfluency is experienced as a significant disability with deleterious effects on functional communication as well as

Correspondence to Claire Cordella: cordella@bu.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

social consequences (Brandenburg et al., 2017; Harmon et al., 2016; Zraick & Boone, 1991). For clinicians, fluency classifications are a crucial source of diagnostic information (Gordon, 1998; Gordon & Clough, 2022). In fact, fluency continues to be an important construct for diagnosis regardless of the etiology of the aphasia. In the poststroke aphasia (PSA) population, umbrella terms of "fluent" or "nonfluent" often serve as a critical step in guiding the decision about specific aphasia subtypes (e.g., Broca's, Wernicke's, anomic, conduction). In the primary progressive aphasia (PPA) population, the three main variants can be at least partially differentiated along a continuum of fluency, from nonfluent (nonfluent variant) to variably fluent (logopenic variant) to fluent (semantic variant; Gorno-Tempini et al., 2011). Given its functional and clinical salience in both poststroke and progressive aphasia, speech fluency is an important construct to assess, monitor, and treat among all PWA.

Speech fluency in aphasia is a multidimensional construct that is determined by various speech and language factors, primarily including motor speech function, grammatical competence, and lexical retrieval ability. This conceptualization has been articulated from theoretical and clinical perspectives (Feyereisen et al., 1991; Goodglass & Kaplan, 1972; Gordon, 1998), and also borne out by data-driven analyses of large numbers of individuals with PSA (Casilio et al., 2019; Clough & Gordon, 2020; Gordon, 2020; Gordon & Clough, 2020; Mirman et al., 2019; Vermeulen et al., 1989). The PPA literature conceptualizes fluency as largely dependent on a combination of syntax and motor speech, as these are the two core criteria for separating out the nonfluent variant (nfvPPA) from the more fluent logopenic (lvPPA) and semantic (svPPA) variants (Gorno-Tempini et al., 2011; Mesulam, 2001). Lexical retrieval is considered to be a main contributor to the variable fluency of logopenic variant PPA. However, in comparison to nfvPPA, this subtype is still typically considered a fluent one owing to spared syntax and motor speech function (Henry & Gorno-Tempini, 2010). Another important nuance relevant to fluency in PPA is that unlike in most cases of PSA, isolated motor speech impairment (i.e., in the absence of aphasia) can be a presenting sign and many, though not all, researchers recognize this presentation with a unique diagnostic designation of primary progressive apraxia of speech (PPAOS; Duffy et al., 2015; Josephs et al., 2012, 2014).

Despite critical advancements in the theoretical understanding of fluency, it remains difficult to measure speech (non)fluency reliably, accurately, and efficiently in PWA. The current clinical gold standard for fluency assessment relies on subjective clinician judgment, via either gestalt judgment or qualitative ratings on standardized fluency scales such as the Western Aphasia Battery–Revised (WAB-R; Kertesz, 2007) or Boston Diagnostic Aphasia Examination (Goodglass et al., 2001). Such metrics of rated fluency require expert assessment and interpretation on the part of trained speech-language pathologists and even then, are prone to unreliability across typical clinician raters (Gordon, 1998; Trupe, 1984). Moreover, the most widely used fluency rating scale for the PSA population—the WAB-R Fluency scale—combines aspects of word-finding, grammatical competence, paraphasias, and motor speech function into a single, pseudoquantitative scale (Casilio et al., 2019). The conflation of each of these different speech and language domains is likely one major reason why the WAB-R fluency scale is virtually unused as a measure of fluency in PPA, wherein grammatical competence and motor speech function dissociate in important diagnostic ways from word-finding and paraphasic errors (Gorno-Tempini et al., 2011; Grossman, 2012; Josephs et al., 2012; Mesulam, 2001; Mesulam et al., 2009). Importantly, more recent attempts at auditory-perceptual ratings of connected speech have demonstrated much improved reliability and the ability to capture a multitude of different aspects of connected speech without conflating them (Casilio et al., 2019). Still, standardized and semistandardized rating scales are inherently limited by their categorical nature; they cannot capture fine-grained nuances in speech fluency in the same way that more continuous quantitative metrics can.

Due in large part to the limitations of existing rating scales for fluency, there has been a decades-long trend toward quantification of fluency in aphasia. Quantitative metrics combine both a numeric value and unit of measurement. For example, a clinician interested in quantitatively assessing speech rate in a connected speech sample may calculate the average rate of speech as words or syllables per second, a quantitative measure. This is in contrast to a more qualitative approach of assessing speech rate as "fast" or "slow" through a gestalt auditory-perceptual decision paradigm. Even when a numerical value is assigned to different categories of auditory-perceptual judgment (e.g., rating speech rate on an operationalized 1–5 scale), these remain inherently qualitative perceptual ratings. Notably, quantification and subjective clinical judgment are not mutually exclusive. As an example, a clinician counting the number of distorted phonemes in a speech sample is making subjective decisions about what constitutes a phonemic distortion; however, the resultant metric combines both a numeric value and unit of measurement and can thus be considered a quantitative measure.

Within the aphasia literature to date, quantification has most typically involved linguistic analyses based on detailed transcription and coding of connected speech for lexical, syntactic, and semantic features (Bastiaanse et al., 1996; MacWhinney et al., 2011; Miller et al., 2015; Prins & Bastiaanse, 2004; Saffran et al., 1989; Thompson et al., 1997; Wilson et al., 2010; Yorkston & Beukelman, 1980).

Quantitative Production Analysis (QPA) was one of the first such quantitative linguistic approaches, and it provided a protocol for transcription and analysis of connected speech in aphasia with a particular focus on syntactic elements (Rochon et al., 2000; Saffran et al., 1989). QPA features are used, for example, to summarize the frequency of occurrence of morphosyntactic elements (e.g., proportion of verbs inflected), complex or grammatical utterances (e.g., proportion of well-formed sentences), or various lexical types (e.g., proportion of closed-class words; Rochon et al., 2000). More recently, approaches to linguistic quantification have focused on automating the calculation of QPA and other features—the most notable example of this type of approach is the Computerized Language Analysis (CLAN) program. CLAN takes as input a detailed manual transcription and uses a variety of built-in programs or customizable command-line analysis options to output quantitative features related to phonology, morphosyntax, syntax, as well as macrostructural discourse measures (MacWhinney, 2018). As with QPA, the connected speech task types from which these measure types are typically derived are semispontaneous in nature, including single picture description, story retell, as well as semistructured interviews (MacWhinney & Wagner, 2010; Stark, 2019). More recently, quantitative approaches have also been focused on more detailed acoustic features as means by which to measure the rate and rhythm of connected speech (Angelopoulou et al., 2018; Cordella et al., 2017, 2019; Feenaughty et al., 2021; Mack et al., 2015; Yunusova et al., 2016), which is understood to be a critical contributor to overall speech fluency (Gordon, 2020; Park et al., 2011; Wagenaar et al., 1975). Acoustic-based approaches have included such measures as subcomponent measures of speech rate (Cordella et al., 2017; Feenaughty et al., 2021) and detailed pause metrics (Angelopoulou et al., 2018; Feenaughty et al., 2021; Mack et al., 2015; Yunusova et al., 2016), including frequency and location of both silent and filled pauses. Similar to linguistic approaches, quantitative acoustic approaches have typically relied on semispontaneous connected speech samples as the basis from which individual acoustic measures are derived. These acoustic features are particularly useful for capturing motor speech aspects of connected speech, which were traditionally neglected in text-based quantitative analysis approaches. Whether analysis is focused on acoustic or linguistic features, there is general optimism that the inclusion of quantitative features can augment clinician judgments of fluency and thereby improve current approaches to fluency assessment in aphasia (Gordon & Clough, 2022).

Although consensus is emerging on the utility of quantitative speech and language measures to complement and inform clinicians' judgments on fluency, there remain several unresolved questions and barriers that hinder clinical adoption of such quantitative approaches. For one, there is limited information available on the comparative time intensiveness of individual measures, in terms of whether and to what degree they require manual transcription or other expert-driven processing (e.g., processing of transcripts through specialized analysis pipelines). This lack of information leads to an assumption that quantitative analysis approaches are uniformly time-consuming, which according to a comprehensive recent survey is likely a major barrier to clinical uptake (Gordon & Clough, 2022). Second, although dozens of individual features have been investigated, there has not been a comprehensive summary of findings across studies to determine which measures are most clinically useful to detect (non)fluency in either PSA or PPA populations. Finally, although fluency is an important criterion for diagnosis of both PSA and PPA, few reports have directly compared fluency in these populations (Ingram et al., 2020), and none to our knowledge none have done so on a meta-analytic level across multiple studies. In practice, clinicians most often assess and treat aphasias of different etiologies and it is therefore important for them to have evidence-based guidance on the appropriateness and transferability of a given assessment or treatment approach for nonfluency. To better facilitate clinical uptake of quantitative approaches for assessing fluency in the future, it is critical to understand from the existing research literature which features are most useful, how these features may be measured, and whether these features are shared across the two main aphasia etiologies.

The purpose of this scoping review is to address these gaps in the current literature and provide a clearer picture of what types of features are being used to assess connected speech fluency in PSA and PPA, as well as the methods used to derive and analyze these features. As part of this review, we also report on the utility of individual quantitative measures to distinguish between fluent and nonfluent PWA, as determined via a summary of the reported statistical significance of each individual feature across all studies. Our review additionally includes a comparative discussion of approaches and features used in PSA versus PPA populations. Finally, we used our findings to offer preliminary guidance to clinicians and researchers who are looking to integrate quantitative measures of speech fluency as part of the assessment process.

Aims of the Current Review

This scoping review sought to characterize quantitative approaches used to assess connected speech fluency in aphasia in recent literature (2012–2022), including both PSA and PPA etiologies. The development of this overarching aim was informed by a combined Population, Phenomenon of Interest, Context (PICo)/Patient, Intervention, Comparison, Outcome, Time (PICOT) framework, which we

then used to generate corresponding search terms (Schardt et al., 2007). The specific aims of the study were to:

1. Characterize individual quantitative features used to index (non)fluency and the transcription and analysis methods used to derive these features.
2. Summarize utility of individual features to differentiate fluent and nonfluent aphasia subgroups.
3. Compare and contrast the types of quantitative features and methods of derivation used in studies of PSA versus PPA.

Method

Protocol and Registration

In conducting this scoping review, we developed an a priori protocol following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Extension for Scoping Reviews guidelines (Tricco et al., 2018). The protocol was prospectively registered on the Open Science Framework on June 30, 2022 (https://osf.io/qxu9y/?view_only=ff58a4ba8de547219fe69ea0bc466f2c).

Study Eligibility

Inclusion/Exclusion Criteria

To be included as part of this review, studies needed to have a focus on connected speech fluency in an acquired aphasia population (PSA or PPA). Peer-reviewed articles were included if they were: (a) published between January 1, 2012, and July 1, 2022; (b) written in English, though language of study participants was unrestricted; (c) reported on at least one quantitative feature (i.e., includes both a numerical value and corresponding unit of measurement) used to measure connected speech fluency or its main contributing components; (d) extracted quantitative feature(s) from a spontaneous or semispontaneous connected speech task; and (e) used quantitative feature(s) to differentiate or characterize (non)fluency using quantitative speech/language features, via a direct-groups comparison of fluent and nonfluent aphasia subgroup(s), correlation analysis between a continuous clinician-rated metric of fluency (e.g., WAB-R Fluency) and a quantitative measure, regression analysis, dimensionality reduction, or machine learning (ML) approach. For the purposes of this review, nonfluent groups included individuals with subtype diagnoses of global, Broca's, transcortical motor aphasia; nfvPPA; individuals with WAB-R Fluency scores of 0–4; individuals labeled as "nonfluent" by the authors; and aphasia + AOS (cf. aphasia alone). Fluent groups included individuals with subtype diagnoses of conduction, anomic, Wernicke's, transcortical sensory aphasia; lvPPA, svPPA; individuals with WAB-R Fluency scores of 5–10; and individuals labeled as "fluent" by the authors. We additionally consider unimpaired healthy controls to be a fluent subgroup. We required specific statistical comparisons to ensure the availability of data that would inform our second research question regarding the utility of individual quantitative features to proxy nonfluency.

Papers were excluded from this review if they (a) were a case study, case series, small cohort study (i.e., reported on < 10 total participants with aphasia or < 5 in any one aphasia subgroup), or review; (b) included only nonquantitative features to index connected speech fluency, such as auditory perceptual rating scales; (c) derived fluency-related quantitative features from a highly controlled experimental task (e.g., sentence completion, read passage, other nonspontaneous speech task, spontaneous speech task) under atypical conditions (e.g., delayed auditory feedback, other perturbation), or from a generative fluency task, as these task types yield nonnatural speech; (d) focused primarily on fluency in a population other than PSA/PPA (e.g., Alzheimer's disease, corticobasal syndrome, stuttering/cluttering, pediatric populations), or in atypical etiologies/presentations of aphasia (e.g., right-hemisphere stroke only, gene mutations, transient aphasia); or (e) focused primarily on phonological processing, morphosyntax, cohesion, coherence, informativeness, communication accommodation, gesture, or other subconstruct tangential to speech fluency. This review was agnostic to the presence of neuroimaging findings related to connected speech fluency; however, studies were excluded if the primary focus of the study was on neuroimaging and the study did not at least include a behavioral-only between-groups statistical comparison in addition to imaging findings.

The literature search was limited to articles published since 2012 due to (a) the focus on modern quantitative approaches, which have evolved significantly in the past decade; and (b) the publication in 2011 of the consensus criteria for the diagnosis of PPA and its main variants (Gorno-Tempini et al., 2011). This was a seminal publication that altered diagnostic practices and terminology for PPA worldwide, and therefore marked a critical shift in categorization and reporting in this population, including within the fluency domain.

Taken together, the inclusion/exclusion criteria for the current study were aimed at identifying modern approaches to the quantification of connected speech fluency in PSA and PPA. Because an aim of the study focuses on the analysis of individual features and their utility to differentiate fluent and nonfluent aphasia groups, we excluded case studies and very small-*N* studies since these either inherently do not include a statistical comparison between a nonfluent and fluent group (case studies) or are liable to be underpowered to detect a between-groups

difference (small-*N* studies) and may thus bias aggregate results regarding the utility of individual features to make this between-groups fluency distinction. Though out of scope of the current study for these reasons, case studies and small-*N* studies nonetheless contribute important information about quantification approaches and features.

Terminological Considerations in PPA

With regard to the PPA literature, we recognize PPAOS as a separate diagnostic entity (cf. nfvPPA; Duffy et al., 2015; Josephs et al., 2012, 2014). Because the focus on this scoping review is fluency in aphasia, we excluded studies focused on PPAOS and its comparison to more fluent, but still grossly nonfluent, subtypes (e.g., progressive apraxia of speech with agrammatism). We acknowledge that due to terminological preferences of individual authors, some patients with isolated AOS may be classified as having nfvPPA, as they technically meet consensus diagnostic criteria requiring either AOS or agrammatism (Gorno-Tempini et al., 2011). Despite this, we nonetheless include for analysis in this review any nfvPPA groups, and make the broad but evidence-based assumption that as a whole, this group of patients is likely to have co-occurring aphasia and AOS (Duffy et al., 2014; Grossman, 2012; Mesulam et al., 2012; Montembeault et al., 2018; Ogar et al., 2007; Tee & Gorno-Tempini, 2019; Vandenberghe, 2016). This decision also makes for a fair comparison to PSA groups included in the review, as these likewise did not include any poststroke pure AOS groups but did include groups with concomitant motor speech impairment reported alongside an aphasia diagnosis. As a final point, we also recognize alternative equivalent or quasi-equivalent terminology in PPA, such as progressive nonfluent aphasia for nfvPPA and semantic dementia for svPPA.

Information Sources and Search Strategy

To identify potentially relevant articles, the following databases were searched systematically: PubMed (U.S. National Library of Medicine); Web of Science Core Collection (Clarivate Analytics); APA PyschInfo (EBSCO); Cumulative Index of Nursing and Allied Health Literature (CINAHL; EBSCO); and Embase (Elsevier). A comprehensive search strategy was developed in consultation with an experienced research librarian and refined through iterative team discussions. We structured search terms around the PICO(T)/PICo question components and generated a list of terms that included main concepts related to the research question: aphasia, nonfluency, connected speech, and quantitative speech and language features (Schardt et al., 2007).

Table 1. Database search terms using a PICO(T)/PICo framework.

Patient, population, or problem	Phenomenon of interest	Context	Outcome	Time
aphasia (poststroke & progressive)	(non)fluency	connected speech	quantitative speech and language features	2012–present
*aphasi**	*fluen**	*spontaneous speech*	*quantitative*	N/A[§]
	*nonfluen**	*connected speech*	*quantif**	
		speech production	*quantitative measure**	
		picture description	*quantitative metric**	
		story	*quantitative feature**	
		narrative	*objective measure**	
		semistructured	*objective metric**	
		discourse	*objective feature**	
			*speech measure**	
			*speech metric**	
			*speech feature**	
			*speech signal**	
			*linguistic measure**	
			*linguistic feature**	
			*lexical measure**	
			*lexical feature**	
			*acoustic**	
			*paus**	
			QPA	
			*automat**	

Note. Italicized text used to denote actual search terms entered. PICO(T) = Patient, Intervention, Comparison, Outcome, Time; PICo = Population, Phenomenon of Interest, Context; N/A = not applicable; QPA = Quantitative Production Analysis; * = truncation wildcard.

[§]Post hoc filter.

As part of our search, we additionally included common variations on these terms (see Table 1). The complete search syntax used for each of the databases is detailed in Supplemental Material S1. Note that the EBSCO databases (APA PyschInfo, CINAHL) were searched jointly using the same syntax. In addition to the systematic keyword searches, we also manually scanned the reference lists of included articles to identify relevant articles that may not have been identified in the systematic searches. Final search results were exported for each database search and uploaded to Covidence, a collaborative web-based review management software (Veritas Health Innovation, 2022). All duplicates were automatically detected across multiple bibliographic imports.

Selection of Sources of Evidence

We used Covidence to implement the two-step screening and selection process for this review:

Step 1: Title/abstract screening. All nonduplicate articles imported into Covidence underwent simultaneous title/abstract screening. Titles and abstracts were screened independently by the first and second authors and evaluated for relevance with reference to the inclusion/exclusion criteria. If it was unclear whether a study met all inclusion criteria or any exclusion criteria, raters erred on the side of including it for subsequent full-text review. Disagreements between the two independent raters were tracked in Covidence and resolved by consensus. Consensus meetings took place approximately 2 weeks after completion of initial reviews and raters were blinded to their initial decisions to include/exclude.

Step 2: Full-text screening. All articles that passed the title/abstract screening stage subsequently underwent full-text review. Full-text articles were again screened independently by the first and second authors and evaluated in reference to the inclusion/exclusion criteria. As before, all inclusion criteria had to be met for a study to be included and only one exclusion criteria had to be met for the study to be excluded. Reasons for exclusion were documented for each excluded article at this stage. Disagreements between the independent raters were tracked in Covidence and resolved by consensus between the two original reviewers. Consensus meetings for this stage took place approximately 1 week after completion of initial reviews, and as before, raters were blinded to their initial decisions. All articles (n = 45) that passed the full-text screening were included in this scoping review.

Data Charting Process

A data extraction template was developed in Covidence by the first and second authors to guide what pieces of critical information to extract from each article. The data extraction template was trialed for a randomly selected subset (n = 5) of included articles. For this subset of articles, the first and second authors jointly extracted data, discussed the results, and made appropriate revisions before publishing a final data extraction template. All remaining included articles were divided between the first and second authors for data extraction.

For each included article, one of the first two authors extracted and charted the following information: (i) etiology of patient population (i.e., PSA; progressive aphasia); (ii) characteristics of included patient groups (i.e., fluent/nonfluent subtype diagnoses, total N of participants, language of study participants, whether/not aphasia severity was reported; whether/not motor speech severity was reported); (iii) method(s) used for derivation of individual quantitative speech/language features (i.e., connected speech task used; degree of automation for transcription and analysis; whether/not transcription time was reported; which automated analysis method was used); and (iv) method(s) used for statistical analysis of quantitative data. Supplemental Material S2 shows the possible categories and detailed options available to raters for charting of items (i)–(iv). In addition to the options listed, each item had an "Other" response option that allowed raters to write in free-form information. Multiselect capabilities were also enabled for all data items. If raters could not find information for any given data item, these were left blank and treated as missing data in subsequent analyses.

For (iv), we categorized the statistical methodologies into the following types determined a priori by the first two authors: group comparisons (e.g., t tests, analyses of variance), correlation analysis (e.g., associating a quantitative attribute with a clinician-rated fluency metric), regression, ML, and dimensionality reduction (e.g., principal component analysis [PCA], factor analysis). We consider ML techniques as distinct statistical models designed to enhance their predictive accuracy through increased data exposure (Miller et al., 2023). This category encompasses straightforward classification strategies that use a single quantitative predictor feature, as well as those that do not implement cross-validation. Cross-validation, a process in which models are trained on specific subsets of available data and tested on separate, independent data subsets, is widely recognized as the ideal method for evaluating the generalizability of an ML model to new data (Bzdok et al., 2017).

It is important to note that we distinguish traditional regression from ML techniques to differentiate between those that primarily aim to infer predictor-outcome relationships and those that use regression primarily to predict an outcome while minimizing prediction error. Similarly, while dimensionality reduction techniques can also form part of ML analysis workflows, we categorize them separately here to highlight studies that utilize dimensionality reduction as a final-stage analysis, rather than as a feature selection technique for ML applications.

For included studies (n = 33) that reported results from a groups difference statistical approach between a nonfluent versus fluent aphasia subgroup(s) or from a correlation approach between a quantitative feature and a clinician-rated fluency metric, see item (iv), above, we charted three additional items per article: (a) individual quantitative speech/language features reported; (b) category of individual speech/language feature; and (c) level of significance of reported result. Items (a) and (c) were charted by either of the first two authors during the extraction process; item (b) was assigned by the first author after all individual features had been extracted and following consultation and iteration on appropriate superordinate categories with other members of the author team. Supplemental Material S3 shows the possible categories and/or assessment schema used by raters for items (a)–(c). For item (a), we recorded the verbatim feature name as reported by authors in the paper. Subsequently, after all raw data for this item had been extracted, the first author undertook a manual renaming process to equate identical and near-identical features (e.g., "WPM" and "speech rate" were renamed to a single "speech rate" feature). This renaming process allowed us to examine patterns and trends in usage of variables over and above simple terminological differences. Importantly, we applied a stringent threshold for what was considered identical or near identical; specifically, we did not consider normalized and nonnormalized feature variants (e.g., # nouns and % nouns) as equivalent because we believed normalization for total quantity of speech output was likely to have an impact on the significance of reported results. We also did not consider features equivalent if they measured the same construct but over different word or sentence contexts (e.g., % pause between sentences vs. % pause within sentences). Likewise, within any given study, all features were treated as nonidentical even if highly similar in deference to the original authors' conceptualization of distinctness. A full list of all verbatim features and manual renaming are shown in Supplemental Material S6, tab: "All Features (PSA + PPA)." For item (b), we recorded the superordinate descriptive category if provided by the authors. We first identified all studies that reported at least two superordinate categories into which individual features were grouped (n = 14 studies; see Supplemental Material S4). From these 14 articles, we extracted, cross-referenced, and reduced these categories to seven broad "constructs," maintaining terminology used by individual author teams wherever possible: quantity, rate/prosody, speech errors, grammatical competence, morphological competence, lexical retrieval ability, and informational content. The author team finalized these seven constructs/categories after discussion and iteration. We then assigned individual features into one of these constructs, again with preference to the original papers' designations, if provided. No feature was allowed to be shared across more than one construct, though we acknowledge that this is a simplification of some more complex quantitative features. The category assignment for each individual feature is listed in Supplemental Material S6, tab: "All Features (PSA + PPA)." These categories are used primarily to discuss individual feature results relative to higher order speech and language constructs in the Discussion section. For Item (c), we recorded the level of significance reported by the authors for every quantitative feature per study, summarized according to a simplified scale (see Table 2). Thus, for every individual feature, we charted a numerical value that captured both the magnitude and direction of significance. This approach was modeled after a recent scoping review (Low et al., 2020) and enabled the comparison of results across studies by individual feature and feature category. The mean significance of a given feature across all studies in which it is reported gives an idea of its utility in differentiating fluent and nonfluent aphasia subgroups, and has the advantage of down-weighting the utility of measures for which results are either null or directly contradictory (e.g., one study shows a significant effect in the direction fluent > nonfluent, another study shows a nonfluent > fluent significant effect).

Critical Appraisal of Individual Sources of Evidence

As this is a scoping review and not a systematic review, we did not formally assess methodological quality or risk of bias as part of a critical appraisal of included articles. We focused instead on providing an overview of

Table 2. Scale for recording reported significance per individual quantitative feature.

Significance of reported result	Description	Assigned numerical value
Significant (positive)	Significant ($p < .05$) group difference (higher nonfluent vs. fluent) OR negative correlation	1
Null	Nonsignificant ($p > .05$) group difference (nonfluent vs. fluent) or correlation	0
Significant (negative)	Significant ($p < .05$) group difference (lower nonfluent vs. fluent) OR positive correlation	–1
Not reported	Significance not assessed and/or reported	–

extant literature as it pertains to our research question, in line with current guidelines for scoping reviews (Tricco et al., 2018). However, for the subset of 33 included studies that reported results from a groups difference statistical approach between a nonfluent versus fluent aphasia subgroup(s)—and from which we extracted and analyzed data about individual speech measures—we did chart additional details about methodological approach. This additional detail includes (a) total *N* of participants with aphasia; (b) whether or not data was used from a shared public data set (e.g., AphasiaBank); (c) whether or not there was significant author overlap among other of the included subset articles (indicating likelihood of overlapping datasets); (d) task type used to derive individual quantitative features; (e) method of transcription and extraction of individual features; (f) *N* or raters/transcribers involved in analysis; and (g) in the event of two or more raters/transcriber, whether or not reliability was reported. The goal of providing this additional level of detail was to offer a critical and transparent accounting of methodological decisions and details that could impact readers' judgments of individual study quality and/or risk of biased results (e.g., resulting from use of same or similar data sets). We also use this additional detail to ground key points in the Discussion section about ways in which future research in this topic area could be made methodologically more rigorous.

Results

Selection and Characteristics of Sources of Evidence

Five hundred and twenty-eight studies were returned based on search criteria across all five databases, with an additional four articles added later based on manual citation searching of included articles. Following automatic removal of 219 duplicate records, 309 total studies underwent title/abstract screening. All 309 titles/abstracts were reviewed independently (by authors C.C. and L.D.) and included or excluded based on prespecified criteria, as detailed in Method section. Interrater agreement for the title/abstract screening stage was 87% and all conflicts were resolved by consensus. A total of 234 studies were excluded following the title/abstract screening process. The remaining 75 studies underwent full-text review, following the same procedure as the title/abstract review (i.e., independent decisions by C.C./L.D., conflicts resolved by consensus). Interrater agreement for the full-text screening stage was 85%. At this stage, specific reason(s) for exclusion were also noted. Forty-five articles survived the full-text screening stage and were included in this scoping review. The full process of identifying and screening sources of evidence is depicted in Figure 1.

For each included article (n = 45), we charted the four primary data items (i.e., etiology, characteristics of patient groups, methods used for derivation of quantitative features, methods used for statistical analysis). Tables 3–6 show the raw charted data, and below, we summarize the key findings for the primary data items.

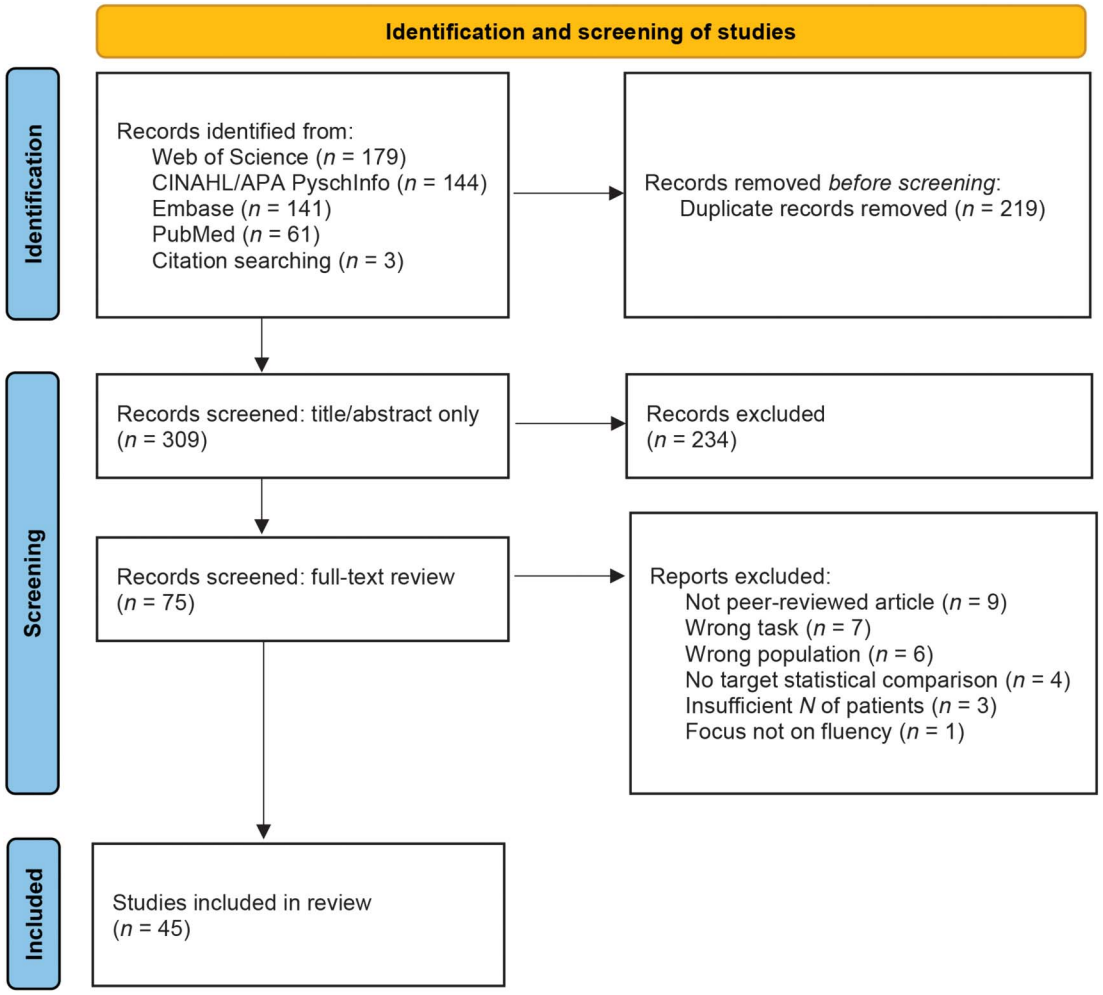
Synthesis of Sources of Evidence

Etiology and Characteristics of Included Patient Groups

Of the 45 included articles, 23 included only patients with PSA (see Table 3), 21 included only patients with PPA (see Table 4), and one study (Ingram et al., 2020) included both PSA and PPA patient populations. Across all studies, the median *N* of the aphasia cohort was 36 (M = 59.9, range: 10–274); median *N* for poststroke studies was 46 (M = 78.4, range: 10–274), compared to a median *N* of 33.5 (M = 39.7, range: 12–88) for progressive studies. The gender distribution (47% female) and age ranges (30–87 years) were consistent with population-level statistics of stroke and aphasia incidence (Reeves et al., 2008). Crucially, although our inclusion/exclusion criteria were agnostic to aphasia chronicity, only two studies included any nonchronic (i.e., acute/subacute) individuals; even in these studies, most patients had chronic aphasia. The language of study participants was likewise unrestricted for purposes of this review, but data were nonetheless heavily skewed toward English (87% of all studies).

Besides basic demographic information, we also charted whether authors reported speech and language severity for included patient groups. The availability of this information was considered important to track, as (non)fluency is likely impacted by overall speech/language severity above and beyond subtype diagnoses and fluent versus nonfluent umbrella groupings. Moreover, because speech and language severity can and do dissociate (e.g., individual with mild language impairment with moderate apraxia of speech), we tracked the reporting of motor speech severity separately from the reporting of language severity. Results revealed that a majority (82%) of studies robustly reported aphasia severity, using either a standardized summary metric (e.g., WAB-R Aphasia Quotient, Clinical Dementia Rating Language subscore) or via detailed reporting of standardized tests to cover a broad range of specific language domains (e.g., syntax, phonological processing, lexical retrieval, etc.). By contrast, only 36% of all studies explicitly reported motor speech severity, using either binary clinician judgments (presence/absence), ordinal ratings (mild/moderate/severe), validated rating scales such as the Apraxia of Speech Rating Scale (Strand et al., 2014), or other robust and well-described approaches. The reporting of motor speech severity was somewhat more consistent among the progressive studies

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis flow diagram detailing identification, screening, and selection of articles for inclusion in the current review.



(41%) compared to the poststroke studies (29%). A summary of these results by etiology is reported in Table 7.

Method(s) Used for Derivation of Individual Speech/Language Features

Of the connected speech tasks used to derive quantitative speech/language features, the most used (n = 23) was a single picture description task (e.g., Cookie Theft), with a picture sequence/story retell task (e.g., Cinderella story) being the second most used (n = 22). Six studies utilized a semistructured interview task. Eight studies reported results for more than one task type.

Besides task type, we also charted subitems related to the transcription and analysis process of connected speech data, as these processes are integral to understanding the derivation of quantitative speech/language features. These results are reported in Figure 2. Note that a single study could use more than one derivation method (e.g., one method for acoustic features, another for text-based feature). For those studies that did include an automation component in their analyses, we also recorded the approach used and did a post hoc grouping to further characterize these methods (see Figure 3). In brief, results showed that a plurality of studies in both the poststroke and progressive literatures derived quantitative speech/language features through a process of manual expert transcription, with subsequent analysis automated. Within this category, the Codes for the Human Analysis of Transcripts to CLAN analysis pipeline was the most used (MacWhinney, 2000, 2018). A sizable percentage of all poststroke and progressive studies included no or minimal automation, relying on manual expert transcription and manual counts or measurements of specific speech/language features. A distinct minority of all studies utilized a more fully automated approach that either did not require expert transcription or automated the transcription

Table 3. Charted data: characteristics of patient groups, poststroke studies.

Study	Nonfluent groups included	Fluent groups included	*N* (aphasia only)	Language of participants	Aphasia severity reported?	Motor speech severity reported?
Alyahya et al., 2020	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia; mixed nonfluent aphasia	Anomic aphasia; conduction aphasia; TCS aphasia; healthy controls	46	English	Detailed subdomain scores	Not reported; Other: excluded pts with "severe motor-speech disorders"
Alyahya et al., 2021	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia; mixed nonfluent aphasia	Anomic aphasia; conduction aphasia; TCS aphasia	46	English	Other: BDAE global severity rating	Not reported; Other: excluded pts with "severe motor-speech disorders as described in the participant's clinical workup"
Clough & Gordon, 2020	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia; WAB-R Fluency 0–5	Anomic aphasia; conduction aphasia; Wernicke's aphasia; TCS aphasia; WAB-R Fluency > 5	254	English	WAB-R AQ	Binary (presence/absence) of AOS & dysarthria
Feenaughty et al., 2021	Broca's/agrammatic aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia	31	English	WAB-R AQ	Binary (presence/absence) of AOS & dysarthria; ASRS
Fromm et al., 2022	Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia; other: not aphasic by WAB-R	168	English	WAB-R AQ	Not reported
Ghoreishi et al., 2020	Global aphasia; Broca's/agrammatic aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia	27	Other: Persian	WAB-R AQ	Not reported
Gleichgerrcht et al., 2021	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia; other: not aphasic by WAB-R	65	English	WAB-R AQ	ASRS; other: included as a control in select stats models not reported as demographic info
Gordon & Clough, 2020	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia; TCS aphasia	254	English	WAB-R AQ	Binary (presence/absence) of AOS & dysarthria
Gordon, 2020	Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia	274	English	WAB-R AQ	Not reported
Gordon & Clough, 2022	Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia	185	English	WAB-R AQ	Binary (presence/absence) of AOS & dysarthria; other: not reported as basic demographic delineated in results instead
Halai et al., 2017b	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia; mixed nonfluent aphasia	Anomic aphasia; Wernicke's aphasia; TCS aphasia	31	English	CDR Language subscore	Not reported

(*table continues*)

Table 3. *(Continued).*

Study	Nonfluent groups included	Fluent groups included	*N* (aphasia only)	Language of participants	Aphasia severity reported?	Motor speech severity reported?
Harmon et al., 2019	Aphasia + AOS	Aphasia only (no AOS, cf. Aphasia + AOS)	14	English	WAB-R AQ	Other: word syllable duration threshold > 330 ms & subjective judgments of sound distortion errors
Ingram et al., 2020	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia; mixed nonfluent aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia; TCS aphasia	106 (76 poststroke)	English	Detailed subdomain scores	Not reported
Kim et al., 2019	Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; conduction aphasia	11	English	WAB-R AQ	Not reported
Kim et al., 2021	Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; conduction aphasia	11	English	WAB-R AQ	Not reported
Kong et al., 2016	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; Wernicke's aphasia; healthy controls	24	English	WAB-R AQ; other: also CLQT language score	Not reported
Kong & Wong, 2018	Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; Wernicke's aphasia; TCS aphasia	68	Other: Cantonese	WAB-R AQ	Not reported
Manning & Franklin, 2016	Author/clinician-defined "nonfluent" aphasia	Author/clinician-defined "fluent" aphasia; healthy controls	22	English	Not reported	Not reported
Martinez-Ferreiro et al., 2017	TCM aphasia; Mixed nonfluent aphasia; other: "motor"	Mixed fluent aphasia; healthy controls	10	Other: Spanish	Other: clinician-defined "mild," "moderate"	Not reported
Mirman et al., 2019	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia	46	English	WAB-R AQ	Binary (presence/absence) of AOS only; other: excluded pts with "significant peripheral dysarthria" AOS based on ABA-2
Nozari & Faroqi-Shah, 2017	Other: authors lump all participants together predict fluency based on continuous WAB-R Fluency scores	Other	112	English	Not reported	Not reported
Zhang et al., 2021	Author/clinician-defined "nonfluent" aphasia	Author/clinician-defined "fluent" aphasia	16	English	Not reported	Not reported; other: dysarthria excluded
Zhao et al., 2020	Global aphasia; Broca's/agrammatic aphasia; TCM aphasia; mixed nonfluent aphasia	Anomic aphasia; conduction aphasia; Wernicke's aphasia; TCS aphasia	70	English	Detailed subdomain scores	Not reported
Zimmerer et al., 2018	Author/clinician-defined "nonfluent" aphasia	Author/clinician-defined "fluent" aphasia; healthy controls	20	English	Not reported	Not reported

Note. TCM = transcortical motor; TCS = transcortical sensory; BDAE = Boston Diagnostic Aphasia Examination; WAB-R AQ = Western Aphasia Battery–Revised Aphasia Quotient; AOS = apraxia of speech; ASRS = Apraxia of Speech Rating Scale; CDR = Clinical Dementia Rating; CLQT = Cognitive Linguistic Quick Test; ABA-2 = Apraxia Battery for Adults–Second Edition.

Table 4. Charted data: characteristics of patient groups, progressive studies.

Study	Nonfluent groups included	Fluent groups included	*N* (aphasia only)	Language of participants	Aphasia severity reported?	Motor speech severity reported?
Ash et al., 2013	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	62	English	Detailed subdomain scores	Not reported; Other: discuss phonetic vs. phonemic errors w/r/t AOS but no dx per pt
Cho et al., 2021	nfvPPA/PNFA	svPPA; Healthy controls	64	English	Detailed subdomain scores	Not reported
Cordella et al., 2017	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	38	English	Other: Progressive Aphasia Severity Scale	Binary (presence/absence) of AOS & dysarthria
Cordella et al., 2019	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	43	English	CDR Language subscore	Ordinal severity rating (mild, moderate, etc.)
Faroqi-Shah et al., 2020	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	26	English	Not reported	Not reported
Fraser et al., 2013	nfvPPA/PNFA	svPPA; Healthy controls	24	English	Not reported	Not reported
Fraser et al., 2014	nfvPPA/PNFA	svPPA; Healthy controls	24	English	Detailed subdomain scores	Other: Those with the nfvPPA had effortful, halting speech with anomia, although not all exhibited clear agrammatism in production or clear apraxia of speech on formal testing
Haley et al., 2021	nfvPPA/PNFA	lvPPA; svPPA	25	English	WAB-R AQ	Ordinal severity rating (mild, moderate, etc.)
Hardy et al., 2016	nfvPPA/PNFA	svPPA; Healthy controls	32	English	Detailed subdomain scores	Not reported
Ingram et al., 2020	nfvPPA/PNFA	lvPPA; svPPA	106 (30 PPA)	English	Detailed subdomain scores	Not reported
Mack et al., 2015	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	35	English	WAB-R AQ	Other: motor speech screening (i.e., oral apraxia screen and repetition of one-, two-, and three-syllable words). nfvPPA group perform worse than other groups on three-syllable words only
Marcotte et al., 2017	nfvPPA/PNFA	svPPA; Healthy controls	25	English	Detailed subdomain scores	Not reported; Other: "nfvPPA patients had effortful speech"
Matias-Guiu et al., 2022	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	76	Other: Spanish	CDR Language subscore	Binary (presence/absence) of AoS & dysarthria
Nevler et al., 2019b	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	59	English	Detailed subdomain scores	Binary (presence/absence) of AoS & dysarthria
Nevler et al., 2020	nfvPPA/PNFA	Healthy controls	23	English	Not reported	Not reported; other: report quant features (e.g., speech rate) that can be proxies for motor speech but no clinician rating of motor speech

(*table continues*)

Table 4. *(Continued).*

Study	Nonfluent groups included	Fluent groups included	*N* (aphasia only)	Language of participants	Aphasia severity reported?	Motor speech severity reported?
Parjane et al., 2021	nfvPPA/PNFA	Healthy controls	25	English	Detailed subdomain scores	Other: medical chart review of subjective characteristics consistent with AOS
Sajjadi et al., 2012	nfvPPA/PNFA	Healthy controls	12	English	Detailed subdomain scores	Not reported
Sitek et al., 2015	nfvPPA/PNFA	lvPPA; svPPA	30	Other: Polish	Not reported	Not reported; other: subjective clinician judgments of speech errors reported as results, not as demo/ characterization info
Themistocleous, Ficek, et al., 2021	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	44	English	CDR Language subscore	Not reported
Themistocleous, Webster, et al., 2021	nfvPPA/PNFA	svPPA; Healthy controls	52	English	CDR Language subscore	Not reported
Thompson et al., 2012	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	37	English	WAB-R AQ	Other: test repeating syllables of varying complexity (out of 50 pts)
Zimmerer et al., 2020	nfvPPA/PNFA	lvPPA; svPPA; Healthy controls	88	English	Detailed subdomain scores	Not reported

Note. nfvPPA = nonfluent variant primary progressive aphasia; PNFA = progressive nonfluent aphasia; lvPPA = logopenic variant primary progressive aphasia; svPPA = semantic variant primary progressive aphasia; dx = diagnosis; pt = patient; AOS = apraxia of speech; WAB-R AQ = Western Aphasia Battery–Revised Aphasia Quotient; CDR = Clinical Dementia Rating.

Table 5. Charted data: methodological approaches to derivation and analysis, poststroke studies.

Study	Method(s) used for statistical analysis of quantitative data	Method(s) used for derivation of individual quantitative speech/language features			
	Statistical approach	Connected speech task used	Degree of automation (transcription + analysis)	Automated analysis method used	Transcription time reported
Alyahya et al., 2020*	Groups difference and/or correlation; dimensionality reduction	Single picture description (e.g., Cookie Theft, Picnic Scene); Picture sequence/story retell (e.g., Cinderella); Other: procedural discourse task	Fully manual transcription & analyses		No
Alyahya et al., 2021*	Groups difference and/or correlation	Single picture description; Picture sequence/story retell; Other: procedural discourse	Fully manual transcription & analyses		No
Clough & Gordon, 2020*	Groups difference and/or correlation; regression	Picture sequence/story retell	Expert transcription (e.g., CHAT, SALT formats), subsequent analyses automated	CLAN	Transcription done not as part of study (e.g., AphasiaBank)
Feenaughty et al., 2021*	Groups difference and/or correlation	Single picture description	Manual (nonexpert) orthographic transcription, subsequent analyses automated; other: semi-automated but requiring manual decisions/processing in Praat	Praat	No
Fromm et al., 2022	Machine learning (ML) and/or classification	Spontaneous free narrative; Single picture description; Picture sequence/story retell; other: procedural discourse	Expert transcription, subsequent analyses automated	CLAN	Transcription done not as part of study
Ghoreishi et al., 2020*	Groups difference and/or correlation	Picture sequence/story retell	Fully manual transcription & analyses		No
Gleichgerrcht et al., 2021	Dimensionality reduction	Single picture description	Expert transcription, subsequent analyses automated	Lu's L2 Syntactic Complexity Analyzer (Python); Stanford POS tagger	No
Gordon & Clough, 2020*	Groups difference and/or correlation; regression	Picture sequence/story retell	Expert transcription, subsequent analyses automated	CLAN	Transcription done not as part of study
Gordon, 2020	Dimensionality reduction; other: groups difference of individual factors	Picture sequence/story retell	Expert transcription, subsequent analyses automated	CLAN	Transcription done not as part of study
Gordon & Clough, 2022*	Groups difference and/or correlation	Picture sequence/story retell	Expert transcription, subsequent analyses automated	CLAN	Transcription done not as part of study

(*table continues*)

Table 5. *(Continued).*

Study	Method(s) used for statistical analysis of quantitative data	Method(s) used for derivation of individual quantitative speech/language features			
	Statistical approach	Connected speech task used	Degree of automation (transcription + analysis)	Automated analysis method used	Transcription time reported
Halai et al., 2017b	Dimensionality reduction	Single picture description	Fully manual transcription & analyses		No, but time mentioned as rationale for features chosen
Harmon et al., 2019*	Groups difference and/or correlation	Picture sequence/story retell	Expert transcription, subsequent analyses automated; fully manual transcription & analyses; other: Praat	Manual transcription + coding, but using Praat	No
Ingram et al., 2020	Dimensionality reduction; ML and/or classification	Single picture description	Other: not specified		No
Kim et al., 2019*	Groups difference and/or correlation	Picture sequence/story retell	Expert transcription, subsequent analyses automated	CLAN	No
Kim et al., 2021*	Groups difference and/or correlation	Picture sequence/story retell	Expert transcription, subsequent analyses automated	CLAN	No
Kong et al., 2016*	Groups difference and/or correlation	Picture sequence/story retell	Fully manual transcription & analyses		No
Kong & Wong, 2018	Regression	Picture sequence/story retell	Fully manual transcription & analyses		No
Manning & Franklin, 2016*	Groups difference and/or correlation	Picture sequence/story retell	Expert transcription, subsequent analyses automated		No
Martinez-Ferreiro et al., 2017*	Groups difference and/or correlation	Other: semistandardized open-ended interview questions (last job, holidays, hobbies)	Fully manual transcription & analyses		No
Mirman et al., 2019*	Groups difference and/or correlation	Picture sequence/story retell	Fully manual transcription & analyses		No
Nozari & Faroqi-Shah, 2017	Other: path modeling approach most similar to linear regression	Picture sequence/story retell	Expert transcription, subsequent analyses automated	CLAN	Transcription done not as part of study
Zhang et al., 2021*	Groups difference and/or correlation	Spontaneous free narrative; single picture description; picture sequence/story retell	Fully manual transcription & analyses		No

(*table continues*)

Table 5. *(Continued).*

Study	Method(s) used for statistical analysis of quantitative data	Method(s) used for derivation of individual quantitative speech/language features			
	Statistical approach	Connected speech task used	Degree of automation (transcription + analysis)	Automated analysis method used	Transcription time reported
Zhao et al., 2020	Dimensionality reduction	Single picture description	Fully manual transcription & analyses		No
Zimmerer et al., 2018*	Groups difference and/or correlation	Other: semistructured interview	Expert transcription, subsequent analyses automated	FLAT	No

Note. CHAT = Codes for the Human Analysis of Transcripts; SALT = Systematic Analysis of Language Transcripts; CLAN = Computerized Language Analysis; POS = part-of-speech; FLAT = Frequency in Language Analysis Tool.

*Study included in subanalysis of individual speech/language features based on availability of appropriate statistical comparison.

Table 6. Charted data: methodological approaches to derivation and analysis, progressive studies.

Study	Method(s) used for statistical analysis of quantitative data	Method(s) used for derivation of individual quantitative speech/language features			
	Statistical approach	Connected speech task used	Degree of automation (transcription + analysis)	Automated analysis method used	Transcription time reported
Ash et al., 2013*	Groups difference and/or correlation	Single picture description (e.g., Cookie Theft, Picnic Scene); Picture sequence/story retell (e.g., Cinderella)	Fully manual transcription & analyses; other: Praat also used, presumably to extract durations	Praat	Yes
Cho et al., 2021*	Groups difference and/or correlation	Single picture description	Expert transcription (e.g., CHAT, SALT formats), subsequent analyses automated	spaCy NLP library (Python)	No
Cordella et al., 2017*	Groups difference and/or correlation; machine learning (ML) and/or classification	Single picture description	Manual (nonexpert) orthographic transcription, subsequent analyses automated	Speech Pause Analysis program (MATLAB)	No
Cordella et al., 2019*	Groups difference and/or correlation; ML and/or classification	Single picture description	Manual (nonexpert) orthographic transcription, subsequent analyses automated	Speech Pause Analysis program (MATLAB)	No
Faroqi-Shah et al., 2020*	Groups difference and/or correlation	Single picture description	Expert transcription, subsequent analyses automated		No
Fraser et al., 2013	ML and/or classification	Picture sequence/story retell	Automated analysis not requiring transcription; expert transcription, subsequent analyses automated	Lu's L2 Syntactic Complexity Analyzer (Python); Stanford POS tagger	No
Fraser et al., 2014*	Groups difference and/or correlation; ML and/or classification	Picture sequence/story retell	Expert transcription, subsequent analyses automated	Lu's L2 Syntactic Complexity Analyzer (Python); Stanford POS tagger	No
Haley et al., 2021*	Groups difference and/or correlation; ML and/or classification	Single picture description; Picture sequence/story retell	Fully manual transcription & analyses; other: Praat also used, but following detailed manual (narrow phonetic) coding	Praat	No
Hardy et al., 2016*	Groups difference and/or correlation	Other: spontaneous propositional speech following structured interview (last holiday)	Other: not enough detail provided		No
Ingram et al., 2020	Dimensionality reduction; ML and/or classification	Single picture description	Other: not specified		No
Mack et al., 2015*	Groups difference and/or correlation; ML and/or classification	Picture sequence/story retell	Expert transcription, subsequent analyses automated; other: also manual coding of pauses	NNLA	No

(*table continues*)

Table 6. *(Continued).*

Study	Method(s) used for statistical analysis of quantitative data	Method(s) used for derivation of individual quantitative speech/language features			
	Statistical approach	Connected speech task used	Degree of automation (transcription + analysis)	Automated analysis method used	Transcription time reported
Marcotte et al., 2017*	Groups difference and/or correlation; dimensionality reduction	Other: topic-directed interviews (what do participants do each day)	Expert transcription, subsequent analyses automated	Lu's L2 Syntactic Complexity Analyzer (Python); Stanford POS tagger	No
Matias-Guiu et al., 2022*	Groups difference and/or correlation; ML and/or classification; dimensionality reduction	Single picture description	Expert transcription, subsequent analyses automated; fully manual transcription & analyses	Praat; SALT software	No
Nevler et al., 2019b*	Groups difference and/or correlation; ML and/or classification	Single picture description	Automated analysis not requiring transcription; fully manual transcription & analyses	In-house (UPenn) SAD	No
Nevler et al., 2020	Groups difference and/or correlation	Single picture description	Automated analysis not requiring transcription	In-house (UPenn) SAD	No
Parjane et al., 2021	Groups difference and/or correlation	Single picture description	Automated analysis not requiring transcription; manual (nonexpert) orthographic transcription, subsequent analyses automated	In-house (UPenn) SAD	No
Sajjadi et al., 2012*	Groups difference and/or correlation	Single picture description; other: semistructured interviews	Fully manual transcription & analyses		No
Sitek et al., 2015*	Groups difference and/or correlation	Single picture description	Fully manual transcription & analyses		No
Themistocleous, Ficek, et al., 2021	ML and/or classification	Single picture description	Fully automated transcription & analyses	In-house speech-to-text "Themis" program (Python); Praat; Textblob library (Python)	No
Themistocleous, Webster, et al., 2021*	Groups difference and/or correlation; ML and/or classification	Single picture description	Fully automated transcription & analyses	In-house speech-to-text "Themis" program (Python); NLTK library (Python)	No
Thompson et al., 2012*	Groups difference and/or correlation	Picture sequence/story retell	Expert transcription, subsequent analyses automated	SALT	No
Zimmerer et al., 2020*	Groups difference and/or correlation; ML and/or classification	Spontaneous free narrative; other: guided personal interview questions	Expert transcription, subsequent analyses automated	FLAT	No

Note. CHAT = Codes for the Human Analysis of Transcripts; SALT = Systematic Analysis of Language Transcripts; NLP = natural language processing; POS = part-of-speech; NNLA = Northwestern Narrative Language Analysis; SAD = Speech Activity Detector; UPenn = University of Pennsylvania; NLTK = Natural Language Toolkit; FLAT = Frequency in Language Analysis Tool.

*Study included in subanalysis of individual speech/language features based on availability of appropriate statistical comparison.

Table 7. Summary of major participant characteristics and task type findings, by etiology.

Summary measure	PSA studies (n = 24)[a]	PPA studies (n = 22)[a]
% studies reporting severity of aphasia	83	82
% studies reporting motor speech severity	29	41
Median *N* of aphasia participants (range)	46 (10–274)	33.5 (12–88)
% studies w/ language of participants = English	88	91
% studies reporting transcription time	0	5
% studies deriving features from . . .		
Single picture description (e.g., Picnic Scene)	38	68
Picture sequence/story retell (e.g., Cinderella)	71	23
Spontaneous free narrative	8	0
Other [examples]	21 [procedural discourse, structured interview]	14 [structured/topic-directed interview]
% studies using > 1 connected speech task	17	18

Note. PSA = poststroke aphasia; PPA = primary progressive aphasia.

[a]Note that the study (Ingram et al., 2020) that included participants with both poststroke and progressive aphasia is double-counted here in both PSA and PPA summary columns.

process. All these studies also featured automated speech/language analysis methods that were either text-based (e.g., natural language processing pipelines) or acoustic (e.g., automatic speech activity detection), though rarely both. These highly automated approaches were more common within the progressive as compared to poststroke literature. This offers a partial explanation for the finding that progressive studies reported a greater mean number of speech/language features per study, since automated methods tend to yield large feature sets and do not impose the same analysis costs (i.e., expert training and availability, time) as more manual methods. Interestingly, the question of analysis cost remains an open question in this body of literature, as only a single study reported the transcription or analysis time associated with their derivation approach.

Figure 2. Summary of transcription and analysis methods used for derivation of individual speech/language features, for poststroke (blue) and progressive (green) studies. Note that a single study could use more than one method type. The *x*-axis is arranged according to degree of automation (blue arrow). Manual t + a = fully manual transcription and analyses; expert t + a = expert transcription (e.g., CHAT, SALT), subsequent analyses automated; nonexpert t + a = manual (nonexpert) orthographic transcription, subsequent analyses automated; auto t + a = fully automated transcription and analyses; auto a only = automated analysis not requiring transcription; other = other or unclear methods that do not fit onto the proposed continuum of automation.

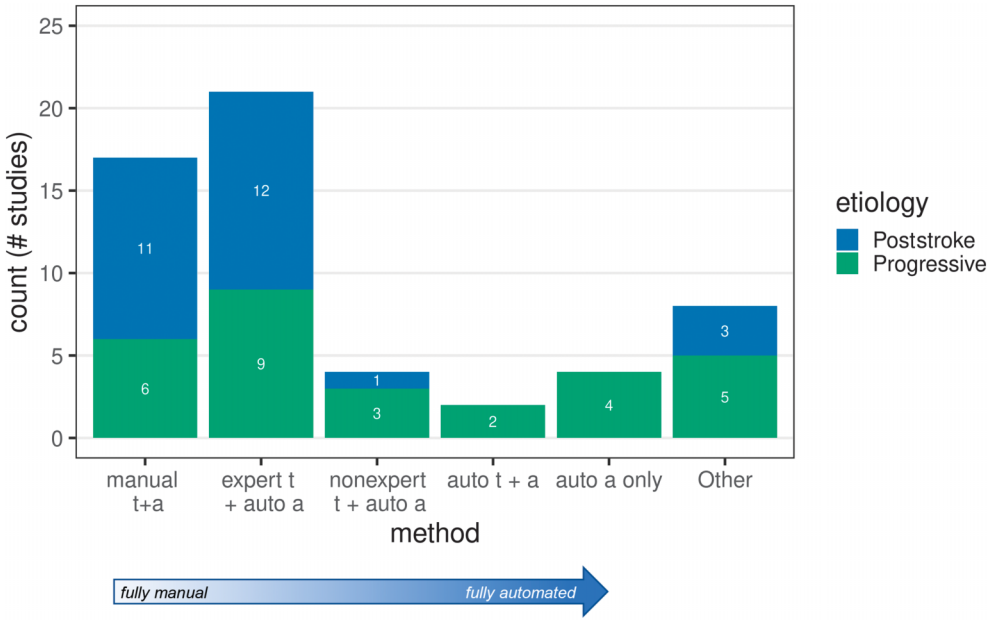
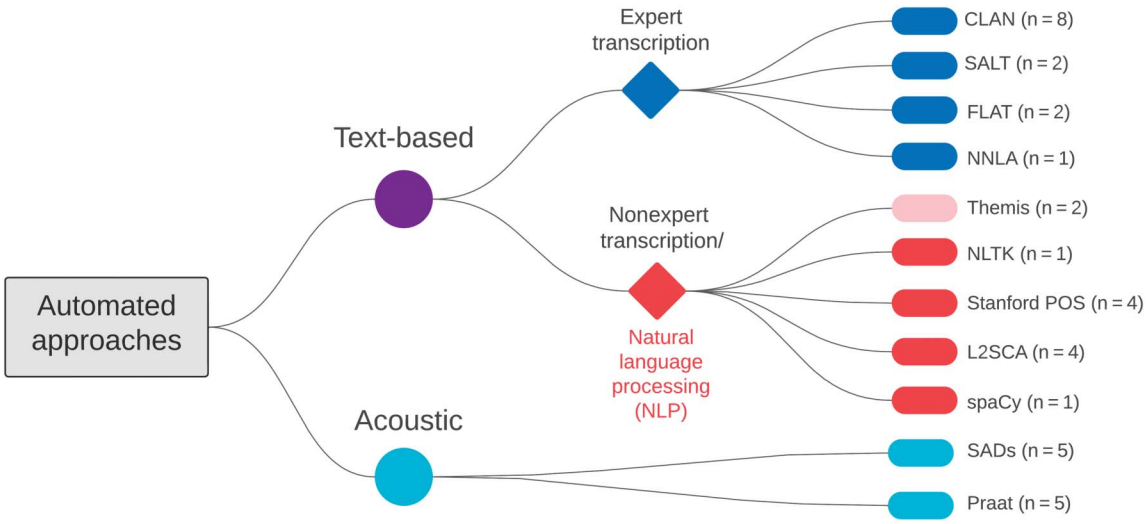


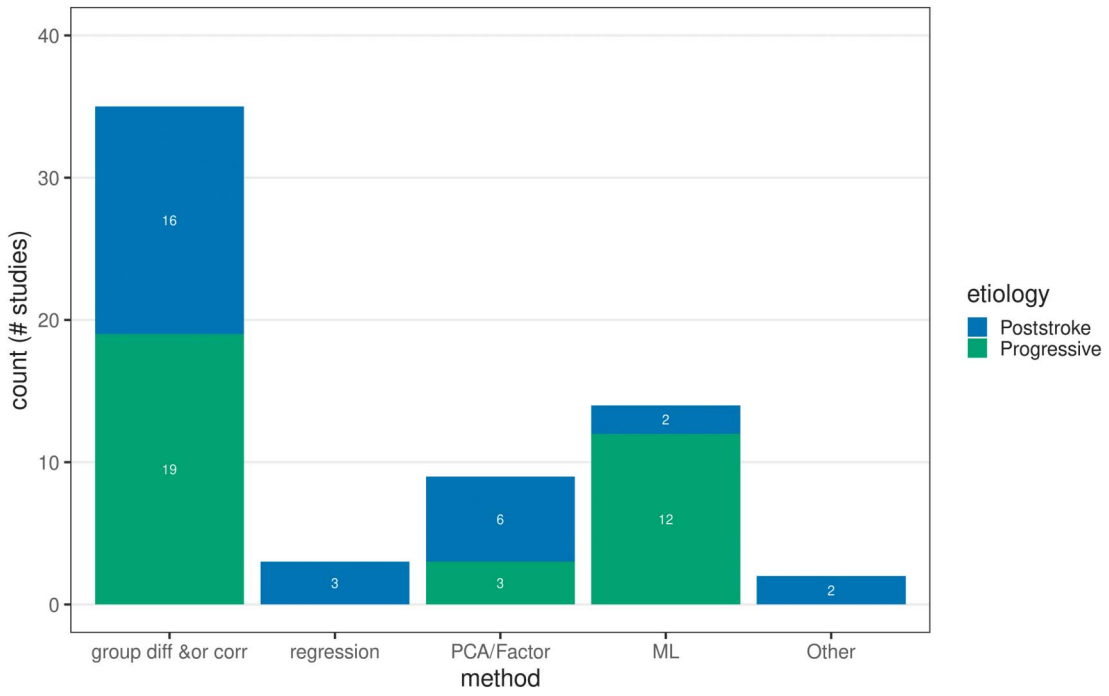
Figure 3. Post hoc categorization of automated approaches used across included studies for transcription and/or derivation of individual quantitative features. CLAN = Computerized Language ANalysis; SALT = Systematic Analysis of Language Transcripts; FLAT = Frequency in Language Analysis Tool; NNLA = Northwestern Narrative Language Analysis; Themis = in-house Python program for automated transcription, which can then be paired with NLP approaches (hence lighter pink shading); NLTK = Natural Language Toolkit; Stanford POS = Stanford Log-Linear Part-of-Speech Tagger; L2SCA = L2 Syntactical Complexity Analyzer; SADs = speech activity detectors.



Method(s) Used for Statistical Analysis of Quantitative Data

For all included studies (n = 45), we charted the general category of statistical approach that was used to differentiate or characterize (non)fluency using quantitative speech/language features. These results are summarized in Figure 4. A marked majority of both poststroke and progressive studies employed a groups difference

Figure 4. Summary of data analysis methods used for poststroke (blue) and progressive (green) studies. Note that a single study could use more than one analysis type. Group diff &or corr = groups comparison (t test, analysis of variance with post hoc) and/or correlation analyses; regression = logistic/linear regression; PCA/Factor = principal components analysis or factor analysis; ML = machine learning (e.g., support vector machine, random forests, simple univariate classification) and/or classification; other = method (e.g., path modeling) not belonging to any of the prespecified categories.



approach wherein one or more nonfluent groups was compared to one or more fluent groups on at least one quantitative feature. Several studies undertook a related approach and correlated a continuous clinician-rated measure of fluency (e.g., WAB-R Fluency subscore) with one or more quantitative features.

Classification and ML analysis methods were less common in the aggregate, though notably among progressive studies, just over half of all studies (12/22) reported these types of predictive results in addition to or instead of null-hypothesis test statistics. This was compared to a smaller minority of poststroke studies (2/25) that did the same. By contrast, dimensionality reduction approaches, and PCA in particular, was more common in poststroke versus progressive literature.

A final noteworthy trend is that the use of ML and dimensionality reduction approaches has increased markedly in the previous 5 years. Looking across both etiologies, only three studies published prior to 2017 employed such an approach, compared to 20 studies published since 2017. This result indicates an increasing trend with time toward more sophisticated statistical approaches.

Individual Speech/Language Features

For a majority subset of included articles (n = 33; denoted with an asterisk in Tables 5 and 6), we charted an additional three data items (i.e., individual quantitative speech/language features reported; category of individual speech/language feature; and level of significance of reported result). This subset was defined as all studies that reported either between-groups significance (specifically nonfluent vs. fluent aphasia subgroups) or correlation results (continuous fluency scale as dependent variable) for one or more individual quantitative speech/language features. This requirement was imposed because we were interested in extracting a significance value per individual speech feature reported that would reflect on the utility of a given measure to differentiate nonfluent and fluency aphasia subgroups or to correlate with an established clinician measure of fluency. Twelve studies did not meet this requirement because they either reported between-groups difference for a nonfluent versus healthy control group only (n = 2) or employed an ML or PCA approach only (n = 10). For these 12 studies, we did record main results per study; however, due to the substantial differences in methodological approach, these results could not be meaningfully summarized in the aggregate and are therefore not included as part of the results to follow.

Methodological details of article subset. For the same subset of articles described above, we also charted specific methodological details to inform readers' judgments of individual study quality. These results are detailed in Table 8. Results show that of the 33-article subset, the mean total N of aphasia participants was 53 (range: 10–274), and a majority of articles (n = 19) reported data from a cohort of at least 30 participants with aphasia. Most articles in the subset (n = 29) reported data from the author group themselves rather than reliance on a shared public database (e.g., AphasiaBank). Importantly, the majority (73%) of the 33-article subset included overlap in first or last authors with at least one other article in the subset. Several of these overlap articles (n = 6) can be explicitly determined to not share data (e.g., focus is on different populations), but this same determination cannot be made for the remaining 18 articles. This relatively high degree of author overlap increases the likelihood that at least some of the primary data across these overlapping articles is the same, thereby introducing a potential bias in results. As in the broader set of all included articles, picture description and story retell tasks were the most commonly used elicitation tasks, and nearly all articles in the subset (29/33) involved some degree of primary transcription in order to derive individual speech/language features. Of these studies, 14 (48%) reported involvement of two or more transcribers/raters, five (17%) reported involvement of one transcriber/rater only, and the remaining 10 (34%) did not report details on the number of transcribers/raters involved. Of the 14 studies reported to involve two or more raters, 10 also reported explicitly on reliability between raters.

Characteristics of speech/language features. Across the 33 articles included in this subanalysis, we extracted and charted a total of 421 individual quantitative speech/language features that were used across studies. After a manual process to equate identical or near-identical features across studies, as described in Method section, results revealed a total of 209 distinct speech/language features across all studies. All features are listed in Supplemental Material S6 for both etiologies, tab: "All Features (PSA + PPA)," and separately by etiology, tabs: "All Features (PSA only)," "All Features (PPA only)." The mean number of speech/language features investigated per study (both etiologies combined) was 28.22 (Mdn = 18; range: 1–58). PPA studies tended to report a greater number of speech/language features: the mean for this group of studies was 34.69 (Mdn = 44; range: 1–58) compared to an M of 9.9 (Mdn = 11; range: 1–16) for PSA studies.

Focusing further on the distinct speech/language features (n = 209), we investigated the degree of overlap of individual features across etiologies (i.e., poststroke vs. progressive). As can be seen in Figure 5 (Panel A), there is minimal overlap in exact features investigated across poststroke and progressive etiologies. In fact, there is just 14% absolute agreement between poststroke and progressive studies across all feature types. Figure 5 (Panel B)

Table 8. Methodological details for 33-article subset.

Study	Study *N* (aphasia only)	Data source	Author overlap*?	What task was used to derive quantitative measures?	How were quantitative features derived?	How many transcribers/ raters were involved?	If multiple transcribers/ raters, was reliability reported?
Alyahya et al., 2020	46	Author data	Yes (Alyahya et al., 2021)	Single picture description (e.g., Cookie Theft, Picnic Scene); Picture sequence/story retell (e.g., Cinderella); Other: procedural discourse task ("how to prepare cup of tea")	Fully manual transcription & analyses	1	N/A
Alyahya et al., 2021	46	Author data	Yes (Alyahya et al., 2020)	Single picture description; Picture sequence/story retell; Other: procedural discourse (how to make tea)	Fully manual transcription & analyses	1	N/A
Ash et al., 2013	62	Author data	Yes (Cho et al., 2021; Nevler et al., 2019b)	Single picture description; Picture sequence/story retell	Fully manual transcription & analyses; Other: Praat also used, presumably to extract durations	2 or more	Not reported
Cho et al., 2021	64	Author data	Yes (Ash et al., 2013; Nevler et al., 2019b)	Single picture description	Expert transcription (e.g., CHAT, SALT formats), subsequent analyses automated	1	N/A
Clough & Gordon, 2020	254	AphasiaBank	Yes (Gordon & Clough, 2020, 2022)	Picture sequence/story retell	Expert transcription, subsequent analyses automated	Transcription done not as part of study (e.g., AphasiaBank)	N/A
Cordella et al., 2017	38	Author data	Yes (Cordella et al., 2019)	Single picture description	Manual (nonexpert) orthographic transcription, subsequent analyses automated	Not reported	Not reported
Cordella et al., 2019	43	Author data	Yes (Cordella et al., 2017)	Single picture description	Manual (nonexpert) orthographic transcription, subsequent analyses automated	1	N/A
Faroqi-Shah et al., 2020	26	DementiaBank	Yes (Themistocleous, Webster, et al., 2021), but different data source	Single picture description	Expert transcription, subsequent analyses automated	Transcription done not as part of study (e.g., AphasiaBank); 2 or more	Yes

(*table continues*)

Table 8. *(Continued).*

Study	Study *N* (aphasia only)	Data source	Author overlap*?	What task was used to derive quantitative measures?	How were quantitative features derived?	How many transcribers/ raters were involved?	If multiple transcribers/ raters, was reliability reported?
Feenaughty et al., 2021	31	Author data		Single picture description	Manual (nonexpert) orthographic transcription, subsequent analyses automated; Other: semi-automated but requiring manual decisions/processing in Praat	Not reported	Yes
Fraser et al., 2014	24	Author data	Yes (Marcotte et al., 2017)	Picture sequence/story retell	Expert transcription, subsequent analyses automated	Not reported	Not reported
Ghoreishi et al., 2020	27	Author data	No	Picture sequence/story retell	Fully manual transcription & analyses	Not reported	Not reported
Gordon & Clough, 2020	254	AphasiaBank	Yes (Clough & Gordon, 2020; Gordon & Clough, 2022)	Picture sequence/story retell	Expert transcription, subsequent analyses automated	Transcription done not as part of study (e.g., AphasiaBank)	N/A
Gordon & Clough, 2022	185	AphasiaBank	Yes (Clough & Gordon, 2020; Gordon & Clough, 2020)	Picture sequence/story retell	Expert transcription, subsequent analyses automated	Transcription done not as part of study (e.g., AphasiaBank)	N/A
Haley et al., 2021	25	Author data	Yes (Harmon et al., 2019), but different populations (PSA vs. PPA)	Single picture description; Picture sequence/story retell	Fully manual transcription & analyses; Other: Praat also used, but following detailed manual (narrow phonetic) coding	2 or more	Yes
Hardy et al., 2016	32	Author data	No	Other: spontaneous propositional speech following structured interview (last holiday)	Other: not enough detail provided	Not reported	Not reported
Harmon et al., 2019	14	Author data	Yes (Haley et al., 2021), but different populations (PSA vs. PPA)	Picture sequence/story retell	Expert transcription, subsequent analyses automated; Fully manual transcription & analyses; Other: Praat	2 or more	Yes
Kim et al., 2019	11	Author data	Yes (Kim et al., 2021)	Picture sequence/story retell	Expert transcription, subsequent analyses automated	2 or more	Yes
Kim et al., 2021	11	Author data	Yes (Kim et al., 2019)	Picture sequence/story retell	Expert transcription, subsequent analyses automated	2 or more	Yes
Kong et al., 2016	24	Author data	No	Picture sequence/story retell	Fully manual transcription & analyses	2 or more	Yes

(*table continues*)

Table 8. *(Continued).*

Study	Study *N* (aphasia only)	Data source	Author overlap*?	What task was used to derive quantitative measures?	How were quantitative features derived?	How many transcribers/ raters were involved?	If multiple transcribers/ raters, was reliability reported?
Mack et al., 2015	35	Author data	Yes (Thompson et al., 2012)	Picture sequence/story retell	Expert transcription, subsequent analyses automated; Other: also manual coding of pauses	2 or more	Not reported
Manning & Franklin, 2016	22	Author data		Picture sequence/story retell	Expert transcription, subsequent analyses automated	2 or more	Yes
Marcotte et al., 2017	25	Author data	Yes (Fraser et al., 2014)	Other: topic-directed interviews (what do participants do each day)	Expert transcription, subsequent analyses automated	Not reported	Not reported
Martinez-Ferreiro et al., 2017	10	Author data	No	Other: semistandardized open-ended interview questions (last job, holidays, hobbies)	Fully manual transcription & analyses	2 or more	Not reported
Matias-Guiu et al., 2022	76	Author data	No	Single picture description	Expert transcription, subsequent analyses automated; fully manual transcription & analyses	1	N/A
Mirman et al., 2019	46	Author data	No	Picture sequence/story retell	Fully manual transcription & analyses	2 or more	Yes
Nevler et al., 2019b	59	Author data	Yes (Ash et al., 2013; Cho et al., 2021)	Single picture description	Automated analysis not requiring transcription; fully manual transcription & analyses	Not reported	Not reported
Sajjadi et al., 2012	12	Author data	No	Single picture description; Other: semistructured interviews	Fully manual transcription & analyses	Not reported	Not reported
Sitek et al., 2015	30	Author data	No	Single picture description	Fully manual transcription & analyses	2 or more	Not reported
Themistocleous, Webster, et al., 2021	52	Author data	Yes (Faroqi-Shah et al., 2020), but different data source	Single picture description	Fully automated transcription & analyses	2 or more	Yes
Thompson et al., 2012	37	Author data	Yes (Mack et al., 2015)	Picture sequence/story retell	Expert transcription, subsequent analyses automated	2 or more	Yes
Zhang et al., 2021	16	AphasiaBank	No	Spontaneous free narrative; Single picture description; Picture sequence/story retell	Fully manual transcription & analyses	2 or more	Yes
Zimmerer et al., 2018	20	Author data	Yes (Zimmerer et al., 2020), but different populations (PSA vs. PPA)	Other: semistructured interview	Expert transcription, subsequent analyses automated	Not reported	Not reported

(*table continues*)

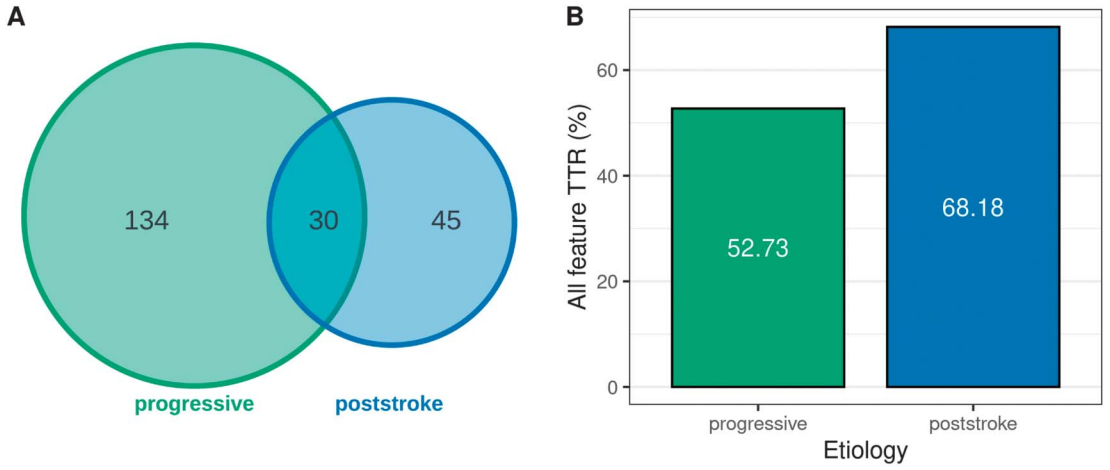
Table 8. *(Continued).*

Study	Study *N* (aphasia only)	Data source	Author overlap*?	What task was used to derive quantitative measures?	How were quantitative features derived?	How many transcribers/ raters were involved?	If multiple transcribers/ raters, was reliability reported?
Zimmerer et al., 2020	88	Author data	Yes (Zimmerer et al., 2018), but different populations (PSA vs. PPA)	Spontaneous free narrative; Other: guided personal interview questions	Expert transcription, subsequent analyses automated	Not reported	Not reported

Note. N/A = not applicable; CHAT = Codes for the Human Analysis of Transcripts; SALT = Systematic Analysis of Language Transcripts; PSA = poststroke aphasia; PPA = primary progressive aphasia.

*Author overlap is defined by a sharing of either the first or one of the last two (senior) authors among any other articles in the 33-article subset.

Figure 5. Featural overlap and agreement within and across etiologies for all studies (n = 33) included in individual feature subanalysis. (A) Venn diagram showing the number of unique speech/language features in the intersection and complements of poststroke and progressive feature sets. (B) Type–token ratio (TTR; calculated as [# distinct individual features / # total individual features] × 100) within each etiology, for all features.



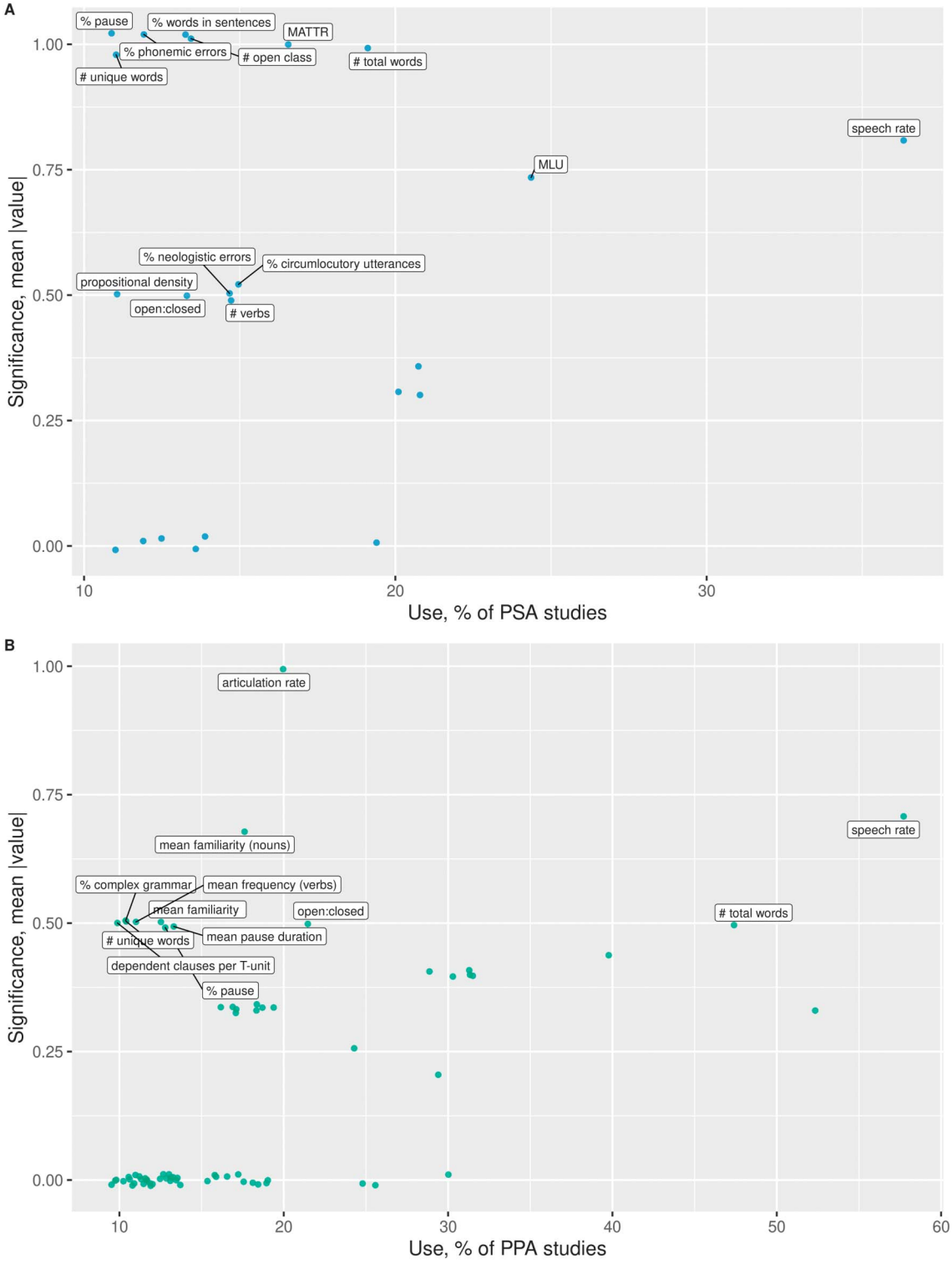
also reports the degree of *within-etiology* variation in the use of specific speech/language features, as measured by a type–token ratio (TTR; i.e., # distinct features; # total features). For example, in a simplified scenario where across all PSA studies, three distinct features were used, with each reported twice (i.e., across two different studies), the TTR would be 3:6, or 50%. A higher TTR is indicative of greater variation in the use of individual features or in other words, a lack of convergence across studies on a core set of features to be investigated. Overall, the all-feature TTR was higher for PSA (68%) compared to PPA studies (53%).

Reported significance of speech/language features. For every individual speech/language feature (n = 421) reported across all studies in this subanalysis, we charted not only the feature itself but also the reported significance. We simplified and discretized reported significance according to a prespecified scale, as described in the Method section (see Table 2), allowing us to summarize these results across all studies in terms of (a) whether or not authors found a significant difference between fluent and nonfluent aphasia subgroups on any particular measure and (b) if significant, what the direction of significance was (i.e., nonfluent < fluent; nonfluent > fluent). Using values from this simplified scale, we calculated the mean value (possible range: −1 to 1) of each individual feature. For this and all subsequent analysis, we analyzed PSA and PPA studies separately for ease of interpretation. Figure 6 shows the side-by-side results of this analysis. Supplemental Material S5 gives corresponding numerical values for these results. For the sake of simplicity, the number of quantitative features listed on the y-axis has been limited to only those (n = 48) that appear at least three times across all studies (PSA and PPA combined). This thresholding lessens the likelihood of overinterpreting the utility of a given feature based on reported significance in only one or a handful of studies. From Figure 6, important differences across the etiologies can be seen. For example, the set of consistently investigated features reported across poststroke studies was smaller in number (27 vs. 46 total features) but included a greater percentage of features with high mean reported significance (i.e., mean absolute value ≥ 0.5). That is, 59% of features in PSA studies met this criterion, compared to just 24% of features in PPA studies. Another important takeaway from Figure 6 is the direction of the significant effect, with darker red indicating a strong and consistent (across studies) significant effect in the direction of nonfluent < fluent performance and darker gray indicating the reverse.

As part of the individual feature analysis, we were also interested in contrasting within etiology the *most used* versus the *most useful* (i.e., most significant, as determined by mean magnitude of significance) features. To this end, Figure 7 displays a scatter plot of features on each of these two dimensions, for PSA (Panel A) and PPA (Panel B) studies. Again, for ease of interpretation the x-axis is thresholded to display only features with use ≥ 10%, and individual features are labeled only if their mean significance value ≥ 0.5. Among PSA studies, three features (speech rate, mean length of utterance [MLU], and # total words) were both (relatively) commonly used and had high mean significance. Among PPA studies, two such features (speech rate, # total words) emerged. Also interesting to note are the features in both etiologies with relatively lower usage rates but high mean significance, as these might be considered potentially promising measures

Figure 6. Mean reported significance level per quantitative feature for all studies (n = 33) included in individual feature subanalysis. Note that figure displays only features occurring at least three times across all studies. "% studies" refers to percentage of studies (within each etiology) reporting significance for a given feature. "Mean Value" refers to average significance value across all studies reporting a given feature. Negative values represent features for which nonfluent < fluent performance (on average), and positive values indicate the reverse relationship. F0 = fundamental frequency; AoA = age of acquisition; TTR = type–token ratio; MATTR = moving average type–token ratio; MLU = mean length of utterance. Color of text on y-axis reflects feature construct: yellow = quantity; green = rate/prosody; red = repairs; orange = speech errors; purple = lexical retrieval ability; green = morphological competence; pink = grammatical competence.

Figure 7. Most used versus most useful features. Scatter plots showing use as measured by the % (within etiology) of studies reporting a given feature by the mean absolute value of the significance of that feature, for (A) PSA and (B) PPA studies. Individual features are labeled only if mean significance value is ≥ .5. *x*-axis is thresholded to display only features with use ≥ 10%. PSA = poststroke aphasia; PPA = primary progressive aphasia; MATTR = moving average type–token ratio; MLU = mean length of utterance.



for future investigation. In PSA, such features include % pause, # unique words, and % words in sentences. In PPA, such features included articulation rate and mean familiarity (nouns). Taken together, results from analysis of individual features highlight the fact that while a large variety of different individual features have been reported across studies, there is considerably more agreement on the most salient features: speech rate and total number of words across both etiologies, along with MLU in PSA studies.

Discussion

Summary of Main Findings

Trends Toward Automated Analysis

As evidenced by the main findings of this review, connected speech analysis is an approach that demonstrates a strong potential for capturing the important dimensions of (non)fluency in aphasia, regardless of etiology. It appears feasible to collect on a large scale, and it is unquestionably possible to extract large amounts of meaningful speech/language features from the resultant samples. Moreover, it is possible—and increasingly common—to automate at least some aspects of the analysis process. In contrast to fully manual methods that were the norm in the early aphasiology literature (Prins & Bastiaanse, 2004; Saffran et al., 1989; Vermeulen et al., 1989; Wagenaar et al., 1975), a majority of all studies in this review used some type of automation to facilitate analysis of connected speech samples. Importantly though, the automation tended to be at the stage of feature extraction and often required as input a detailed manual transcription done by trained experts. A handful of reviewed studies automated or simplified transcription, for instance by generating a nonexpert orthographic transcription and pairing this with a natural language processing analysis approach (Fraser et al., 2013; Themistocleous, Ficek, et al., 2021; Themistocleous, Webster, et al., 2021). Still other studies employed a text-free or text-minimal acoustic approach, although this approach does limit the breadth of feature types that can be investigated (Cordella et al., 2019; Nevler et al., 2019b, 2020).

Emergence of Advanced Statistical Approaches

Review results demonstrated that although most reviewed studies continue to focus primarily on between-groups differences (i.e., fluent vs. nonfluent) and inferential statistics, an increasing number of studies (n = 20) published since 2017 have employed ML or other advanced statistical approaches. This is compared to only three such studies employing these types of analysis approaches in the period from 2012–2016. This trend observed in the current review mirrors broader trends reported in the general aphasia literature on the increasing use of ML in recent years (Adikari et al., 2023). A potential advantage of ML approaches is the ability to capture nonlinear relationships between large numbers of input features and a desired output and use this information to inform future clinical decision making. Particularly when these models are cross-validated with different sets of training and test data, they may enhance the generalizability of results and make it more likely that features identified as important for determining nonfluency in one study could be applied to more varied contexts.

If we take as our goal—as has been well-articulated by others (see, for example, Gordon & Clough, 2022)—to be able to reliably quantify and characterize an individual's fluency status as part of the fuller characterization of their clinical profile, then we need to understand the dimensions and features that are most meaningful to clinical characterizations. ML and data reduction techniques are promising approaches for the identification of a smaller set of core speech/language features that might be most important or meaningful for characterizing nonfluency in aphasia. Reducing feature redundancy is a critically important endeavor, especially given the time intensiveness of current analysis methods (Armstrong et al., 2007; Bryant et al., 2016).

Variation Across Studies in Specific Measures Used to Index Nonfluency

Review results revealed a large number of distinct individual speech/language features used across and within PSA and PPA etiologies, leading to low levels of absolute cross-etiological agreement (i.e., < 15% of the same individual features used across PSA and PPA studies) and relatively high variation within-etiology. Review results showed that for all features (i.e., combined across all categories), the TTR was 68% for PSA studies and 53% for PPA studies. This means that, particularly in the PSA literature, there is a tendency to use a wide variety of different individual features, which naturally leads to low levels of convergence on these features across studies. Importantly though, the vast majority of these individual features can be coherently categorized into a smaller number of underlying constructs, which are concordant with the pillars of nonfluency, namely lexical retrieval, motor speech, and agrammatism in PSA and motor speech and agrammatism in PPA.

Primacy of Quantity, Rate, and Syntactic Features for Identifying Nonfluency

Among the large number of candidate features reported in the literature, review results demonstrated that the most useful features for differentiating fluent from nonfluent aphasia in both the PSA and PPA literature—based on the mean magnitude of significance—were

features related to overall quantity of speech, rate of speech, or syntactic complexity and/or accuracy. Lexical variables related to lexical content, diversity, and accuracy were also well-represented among highly significant features but the direction of these effects (i.e., whether the given feature was higher or lower in nonfluent compared to fluent aphasia) was discrepant across etiologies and may highlight an important difference in the presentation of nonfluency in PSA versus PPA, a point we discuss in further detail below.

Within the PSA literature summarized as part of this review, quantity (# total words) or lexical content features confounded with quantity (# unique words, # open class, # verbs), rate (speech rate, % pause), and syntactic features (% words in sentences, MLU) were among the most useful for identifying nonfluency in this population (as evidenced by relatively high mean significance value in Figure 7A), Specifically, results demonstrated that across studies, nonfluency was robustly associated with reduced total quantity of speech, slowed rate of speech at least partially attributable to increased pausing, and reduced grammatical complexity of connected speech. Importantly, the finding for primacy of speech quantity in particular in determining connected speech fluency in PSA converges with very recent evidence from an unsupervised, data-driven analysis using connected speech to identify naturally occurring clusters in a large, heterogeneous PSA cohort (Fromm et al., 2022). In this analysis, total number of words and total number of closed class words were the only features needed to identify clinically meaningful clusters. With regard to lexical variables, review results revealed lexical diversity—as measured by the moving average type–token ratio—to be among the most robust differentiators between fluency subgroups, with studies consistently reporting significantly reduced lexical diversity for nonfluent compared to fluent PSA subgroups. Somewhat less consistently, studies also reported a greater percentage of neologistic errors, reduced propositional density but contrastingly, a lower percentage of circumlocutory utterances for nonfluent (cf. fluent) subgroups (see Figure 7A). Taken together, review results support existing conceptualizations of nonfluency in PSA as multiply determined by motor speech, grammatical, and lexical retrieval deficits.

Within the PPA literature summarized as part of this review, quantity (# total words) or lexical content features confounded with quantity (# unique words), rate (articulation rate, speech rate, % pause), and syntactic features (% complex grammar, dependent clauses per T-unit) were among the most useful for identifying nonfluency in this population (see Figure 7B). Similar to PSA, results demonstrated that across PPA studies, nonfluency was strongly associated with reduced total quantity of speech, slowed rate of speech attributable to increased pausing and reduced articulation rate, as well as reduced grammatical complexity of connected speech. The top two most robustly significant variables across PPA studies were speech rate and articulation rate, the latter of which is a subcomponent rate measure that excludes pause and has been attributed to motor speech function in PPA (Cordella et al., 2017). This finding for the salience of rate measures in PPA is strikingly consistent with recent results from a predictive, ML analysis in which multiple algorithms ranked a speech rate measure as the single most important feature for subtype classification (Matias-Guiu et al., 2022). Somewhat in contrast to PSA findings, lexical variables demonstrated no reliably significant differences in lexical diversity between fluent and nonfluent groups; moreover, the nonfluent variant tended to use less frequent and less familiar individual words compared to more fluent PPA subtypes. These findings highlight a unique aspect of fluency and its conceptualization in PPA as compared to PSA—namely that lexical deficits have never been considered to be a defining feature of the nonfluent variant of PPA (Gorno-Tempini et al., 2011; Grossman, 2012; Ogar et al., 2007). Taken together, review results for PPA studies are consistent with existing conceptualizations of nonfluency in PPA as principally determined by underlying motor speech impairment and agrammatism, the two core criteria for diagnosis of nfvPPA (Gorno-Tempini et al., 2011). That is, the syntactic features that proved useful in identifying nfvPPA across reviewed studies can be interpreted to support a degree of agrammatism in this group compared to more fluent subtypes. Somewhat more obliquely, both rate and quantity measures can be mapped to motor speech function, including in PPA (Ingram et al., 2020; Poole et al., 2017). For example, reduced speech rate is one of the most commonly cited diagnostic features of acquired apraxia of speech according to a recent review (Allison et al., 2020). Moreover, subcomponents of speech rate such as articulation rate or maximum phonation rate—which reflect only the rate of spoken syllables or words, disregarding pause time—have been associated specifically with clinical presentations of motor speech dysfunction and/or speech motor-involved brain regions in PPA and related disorders (Cordella et al., 2019, 2022; Duffy et al., 2017; García et al., 2022; Josephs et al., 2023).

The Fluency Construct in Poststroke and Progressive Etiologies

A major goal of this review was to enable meaningful comparison of fluency and the ways in which it is measured across the two major acquired aphasia etiologies. In both PSA and PPA, fluency is a key construct with important diagnostic and therapeutic implications. There is debate, however, as to whether fluency is a comparable construct across these two etiologies and therefore whether it can or should be quantified in similar ways (Ingram et al., 2020).

One the one hand, there is indirect evidence that the constructs subserving speech fluency are similar across PSA and PPA etiologies albeit with a few important distinctions. As previously mentioned, the PSA literature conceptualizes fluency as multiply determined by three main constructs: syntax, motor speech, and lexical retrieval (Goodglass & Kaplan, 1972; Gordon, 1998, 2020). The progressive literature conceptualizes fluency as largely dependent on a combination of syntax and motor speech (Gorno-Tempini et al., 2011; Mesulam, 2001; Mesulam et al., 2012). Review results showed specifically that features related to quantity, rate, and syntax are both commonly used and show robust overall significance in differentiating fluent from nonfluent subgroups in both PSA and PPA. Lexical features are useful for differentiating fluency subgroups in PSA, with nonfluency associated with impoverished lexical access. By contrast, nonfluency in PPA is not associated with impoverished lexical content, at least on lexical metrics that are not confounded with overall quantity.

On the other hand—despite some shared underlying constructs to fluency—there is limited but compelling evidence of etiological differences in the manifestation of fluency when assessed on the same scale. In one of very few studies to include both PSA and PPA etiologies in a single data-driven analysis, Ingram and colleagues found that unlike other linguistic dimensions (e.g., semantics, phonology), speech fluency revealed a strong etiological separation whereby even so-called fluent PSA subtypes were less fluent than virtually all PPA subtypes (Ingram et al., 2020). The current review adds to this line of inquiry not by directly comparing performance on any one quantitative feature, but by surveying the individual features used and the strength of evidence per feature within each etiology. Overall, results suggest a tendency to use nonidentical individual features across etiologies. However, results also demonstrated important similarities across etiologies. That is, when considering the features within each etiology that showed the greatest mean significance, there was considerable overlap across PSA and PPA studies, suggesting a degree of featural overlap among higher saliency features. It may thus be possible to measure fluency across etiologies with similar features or feature sets, even if we acknowledge that performance on those measures will likely differ markedly, for example, in comparing nonfluent PSA subtypes to nfvPPA.

It should be noted that in both etiologies, there is a push away from binary or rigid categorical conceptualizations of fluency, with many compelling arguments that these categories are unreliable to diagnose, often overlapping, and/or hold minimal predictive value for recovery or decline (Casilio et al., 2019; Ingram et al., 2020; Wilson et al., 2022). Instead, many of these arguments advocate for thinking about fluency as a graded, continuous dimension as part of a multidimensional profiling of speech and language deficits. In these conceptualizations, fluency remains an important construct to assess and treat; in fact, several data-driven studies have returned fluency as a primary dimension explaining overall variance in connected speech output in large aphasia cohorts (Alyahya et al., 2020; Fromm et al., 2022; Halai et al., 2017a; Matias-Guiu et al., 2022). These newer conceptualizations do, however, necessitate a change in how fluency is measured. Specifically, they require a shift away from coarse subjective categories and toward more reliable, finer grained measurement of fluency. Quantification of fluency from connected speech samples is one such approach that is promising in this regard. Results from this review have provided a snapshot of current approaches to the quantification of connected speech fluency in aphasia. Below, we discuss the implications of these findings as applied to the clinical setting and explore potential ways forward in pursuit of reliable, feasible, and interpretable quantification approaches.

Clinical Implications, Barriers, and Future Directions

As evidenced by the results of this review, quantitative assessment of connected speech shows promise for fine-grained characterization of fluency in aphasia. Yet, despite its widespread use among researchers, subjective categorical or semicategorical ratings of fluency remain the clinical gold standard. Data from this scoping review can shed light on some of the potential barriers to clinical uptake of current quantification approaches, as well as highlight promising trends and approaches for future use.

Barriers to Clinical Uptake

Current approaches to quantification of speech fluency in aphasia are limited by two critical realities as revealed in this review: (a) features proposed to date overwhelmingly rely on time-consuming manual derivation methods; and (b) there is little consensus on which, how many, or what type of quantitative features ought to be used as proxies for fluency. An additional caution of the literature centers on potential methodological shortcomings and/or biases of current approaches.

Regarding barrier (a), results of this review demonstrated a continued reliance on at least partially manual feature extraction methods, particularly for the transcription stage of analysis. As has been articulated in other reviews of the broader discourse literature (Bryant et al., 2016, 2017; Stark et al., 2021), manual transcription is a significant barrier to implementation within clinical settings, where time is often the limiting factor. Similarly, a recent survey of current clinical practice for assessing speech

fluency in aphasia found that a majority of SLPs recognized a need for more objective, fine-grained assessment methods but were reluctant or unable to use time-intensive measures in everyday clinical practice (Gordon & Clough, 2022). In the current review, we considered a narrow aspect of time intensiveness, namely whether or not authors reported transcription time. Our findings indicated that virtually no studies reported the approximate per sample time required for transcription. This means it is not possible to fully evaluate transcription and analysis methods in terms of their time cost and suggests a significant gap in the communication of important information to a clinical audience.

Regarding barrier (b), review results revealed 210 distinct candidate features that might possibly be useful for differentiating nonfluent and fluent aphasia subtypes. Even within etiology, there was a tendency across studies to investigate related but nonidentical features as proxies for the same core construct. Though this exploration of large numbers of potential features is useful from a research perspective, it must also be accompanied with approaches that critically evaluate the utility of proposed features. A clinician wishing to undertake a quantitative analysis needs to know the core set of nonredundant features that capture the most important aspects of speech fluency, and that information is currently lacking in the literature to date. Identification of a smaller subset of crucial features—combined with a detailed explanation of what each of these features represent and how best to measure them—would make for a more coherent message that could more easily be adopted into the clinical context. Current review results may be a helpful starting point toward this end by identifying features that are most used as well as those that are most useful in differentiating fluency subgroups across a sizable number of studies in both the PSA and PPA literature. Going forward, the best approach to identifying an optimal feature set for clinical adoption may be advanced statistical and ML approaches. We discuss more on this point in the following section.

A final limitation of the current literature relates to methodological details of the studies used to extract individual speech/language features identified and analyzed in this review. Although, in keeping with the standards of scoping review, we did not include a formal assessment of methodological quality of included articles, we nonetheless summarized key methodological details of all articles from which individual speech/language features were extracted. Results on this point revealed a substantial risk in redundant primary data across included subset articles, stemming largely from the fact that the same or similar author groups published multiple studies reporting on potentially some of the same participants. Several other studies in the article subset make use of a shared public database (either AphasiaBank or DementiaBank), which also means results are based on the same or similar set of individual participants. Although important to consider in interpreting individual study results, we do not view these as inherent weaknesses and some degree of data overlap is unavoidable, particularly when dealing with rare disorders. In addition, shared databases such as AphasiaBank have the advantage of data being drawn from a diverse set of nationwide centers, with resulting data transcription subject to a highly controlled, rigorous, and transparent process (MacWhinney et al., 2011). The other key methodological point revealed by our charting process is the inconsistent approach to manual transcription and analysis. Crucially, studies did not reliably report involvement of two or more transcribers/raters and formal reliability between raters was reported even more rarely. Future research aiming to extract individual speech/language measures from connected speech would likely be strengthened by inclusion of multiple raters and reliability analyses, as these safeguards would increase confidence in the measures themselves. This type of enhanced methodological reporting could also serve to identify measures that, while promising in their ability to differentiate fluency subgroups, might also be unreliable to extract and therefore not suitable for use in the clinical context.

A Multipronged Path Forward

Understanding the limitations of the literature, there are nonetheless promising advances underway that are likely to improve current approaches to assessing speech fluency in aphasia. Regarding quantification of connected speech, recent research has introduced increasingly efficient preprocessing approaches, including the use of natural language processing or acoustic-based analysis techniques that could lessen or eliminate the need for detailed expert transcriptions (Cho et al., 2021; Fraser et al., 2013; Liu et al., 2023; Nevler et al., 2019a; Themistocleous, Ficek, et al., 2021; Themistocleous, Webster, et al., 2021). There are also major advances being made to automate transcription for individuals with aphasia or other speech and language disorders (Gosztolya et al., 2019; Jacks et al., 2019; Themistocleous, Ficek, et al., 2021; Themistocleous, Webster, et al., 2021). These types of approaches have the potential to enhance the utility of connected speech analysis by enabling efficient processing and analysis of large volumes of connected speech data collected in everyday, ecologically valid settings (e.g., from audio recordings of standard in-clinic neuropsychological examinations; at home via smartphone app). Such approaches are already being used to promising effect among other patient populations (Amini et al., 2022; Connaghan et al., 2019; Stegmann et al., 2020; Tavabi et al., 2022; Xue et al., 2021).

Simultaneously, there are recent trends toward utilization of advanced predictive modeling and data-driven

approaches, including ML (Adikari et al., 2023). These types of approaches enable researchers to develop models that take as inputs large numbers of multidimensional features and reveal pared down feature sets with optimal explanatory power and/or diagnostic relevance. To date, ML approaches in the aphasia literature have typically involved supervised classification or regression tasks that map various hand-crafted input variables (e.g., quantitative speech/language features, demographic variables, imaging-based measures) to clinically labeled output variables (e.g., diagnostic subgroup, treatment response; Billot et al., 2022; Bonilha et al., 2019; Kristinsson et al., 2021). More recently, studies have begun to combine supervised and unsupervised approaches, in acknowledgment of the often-imperfect clinical labels used as ground truth in supervised-only approaches. Fromm and colleagues used one such combined supervised/unsupervised approach to identify connected speech predictors of naturally occurring aphasia subtype clusters, reducing a set of 221 input features to just two critically important ones (Fromm et al., 2022). In this way, ML-based approaches can bring researchers closer to convergence on a particular set of clinically useful features. Importantly though, ML approaches must continue to be rigorously evaluated for accuracy, risk of bias in predictions, and clinical interpretability of results (Char et al., 2018; Grollemund et al., 2019; Yoon et al., 2022).

The final bigger picture consideration for the future is recognizing the appropriate *role* for quantification approaches in the clinic setting. At least for the foreseeable future, quantification approaches are unlikely to supplant clinician ratings. A more realistic short-term goal—and one that is already underway (Casilio et al., 2019; Gordon & Clough, 2022)—is the development of more reliable and fine-grained subjective scales that incorporate quantification where appropriate. As an example, notoriously subjective ratings of motor speech severity might be augmented by quantifying prosodic (e.g., speech rate) and phonetic (e.g., error counts) aspects of connected speech (Jacks et al., 2019). Importantly, using quantification in this way still necessitates an understanding of which features are the best proxies for the different contributing aspects of fluency. It also requires time-efficient methods to extract these features. For these reasons, continued advances toward efficient, reliable quantification of connected speech fluency remains a critical endeavor for aphasia researchers.

Limitations

There are several limitations of the current review. First, we restricted our literature search to articles published since 2012. Although we believe this delimiter to be crucial in allowing us to compare results more meaningfully across PSA and PPA etiologies, it does mean the review offers only a snapshot of recent trends. This is important to keep in mind, particularly regarding the PSA literature, which has a longer history compared to the PPA literature. One practical consequence of this may be that the PSA literature is now less reliant on categorical comparisons and binary classification (fluent vs. nonfluent) of fluency, whereas this categorical conceptualization is still embedded into the current PPA consensus criteria and therefore continues to be referenced quite widely in that literature. Because our review focused primarily on studies that made a direct comparison in some form between fluent and nonfluent aphasia subtypes, our results may reflect an overrepresentation of progressive studies if indeed these types of comparisons are currently more common than in the PSA literature.

A second limitation of the current study is the reliance on subjective judgment to equate and categorize individual features extracted from the included studies. Many individual features (e.g., # utterances, speech rate) could plausibly be categorized into more than one superordinate category, and yet, for the sake of simplicity, we imposed a one-to-one relationship between feature and category. Moreover, we used a stringent method for equating identical or near-identical features. This approach means that our results may underestimate the degree of convergence—both across and within etiologies—if evaluated more holistically.

A final limitation of this study relates to the representation of both the same participants and same authors across multiple of the included studies. The representation of the same participants—notably via mining of large, shared databases (e.g., AphasiaBank)—may have implications for interpreting the significance of variables in distinguishing between fluent and nonfluent PWA. The representation of the same authors across multiple studies presents a similar issue: authors may use overlapping participant data sets across different studies and also tend to use similar methodological approaches and/or feature sets, which may bias the representation of certain features in our review.

Data Availability Statement

Data used for the current analysis will be made available upon reasonable request by any qualified investigator.

Acknowledgments

This work was funded by National Institutes of Health Grant F32DC020342 (awarded to C.C.). In addition to the funding sponsor, the authors wish to thank Kate Silfen, Health Sciences Librarian at Boston

University, for her invaluable help with the development of search terms and iterative feedback on database search syntax.

References

*Indicates inclusion in subanalysis of individual quantitative features.

Adikari, A., Hernandez, N., Alahakoon, D., Rose, M. L., & Pierce, J. E. (2023). From concept to practice: A scoping review of the application of AI to aphasia diagnosis and management. *Disability and Rehabilitation*. Advance online publication. https://doi.org/10.1080/09638288.2023.2199463

Allison, K. M., Cordella, C., Iuzzini-Seigel, J., & Green, J. R. (2020). Differential diagnosis of apraxia of speech in children and adults: A scoping review. *Journal of Speech, Language, and Hearing Research, 63*(9), 2952–2994. https://doi.org/10.1044/2020_JSLHR-20-00061

***Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A.** (2020). A unified model of post-stroke language deficits including discourse production and their neural correlates. *Brain, 143*(5), 1541–1554. https://doi.org/10.1093/brain/awaa074

***Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A.** (2021). Content word production during discourse in aphasia: Deficits in word quantity, not lexical–semantic complexity. *Journal of Cognitive Neuroscience, 33*(12), 2494–2511. https://doi.org/10.1162/jocn_a_01772

Amini, S., Hao, B., Zhang, L., Song, M., Gupta, A., Karjadi, C., Kolachalama, V. B., Au, R., & Paschalidis, I. C. (2022). Automated detection of mild cognitive impairment and dementia from voice recordings: A natural language processing approach. *Alzheimer's & Dementia, 19*(3), 946–955. https://doi.org/10.1002/alz.12721

Angelopoulou, G., Kasselimis, D., Makrydakis, G., Varkanitsa, M., Roussos, P., Goutsos, D., Evdokimidis, I., & Potagas, C. (2018). Silent pauses in aphasia. *Neuropsychologia, 114,* 41–49. https://doi.org/10.1016/j.neuropsychologia.2018.04.006

Armstrong, L., Brady, M., Mackenzie, C., & Norrie, J. (2007). Transcription-less analysis of aphasic discourse: A clinician's dream or a possibility? *Aphasiology, 21*(3–4), 355–374. https://doi.org/10.1080/02687030600911310

***Ash, S., Evans, E., O'Shea, J., Powers, J., Boller, A., Weinberg, D., Haley, J., McMillan, C., Irwin, D. J., Rascovsky, K., & Grossman, M.** (2013). Differentiating primary progressive aphasias in a brief sample of connected speech. *Neurology, 81*(4), 329–336. https://doi.org/10.1212/WNL.0b013e31829c5d0e

Bastiaanse, R., Edwards, S., & Kiss, K. (1996). Fluent aphasia in three languages: Aspects of spontaneous speech. *Aphasiology, 10*(6), 561–575. https://doi.org/10.1080/02687039608248437

Billot, A., Lai, S., Varkanitsa, M., Braun, E. J., Rapp, B., Parrish, T. B., Higgins, J., Kurani, A. S., Caplan, D., Thompson, C. K., Ishwar, P., Betke, M., & Kiran, S. (2022). Multimodal neural and behavioral data predict response to rehabilitation in chronic poststroke aphasia. *Stroke, 53*(5), 1606–1614. https://doi.org/10.1161/STROKEAHA.121.036749

Bonilha, L., Hillis, A. E., Wilmskoetter, J., Hickok, G., Basilakos, A., Munsell, B., Rorden, C., & Fridriksson, J. (2019). Neural structures supporting spontaneous and assisted (entrained) speech fluency. *Brain, 142*(12), 3951–3962. https://doi.org/10.1093/brain/awz309

Brandenburg, C., Worrall, L., Copland, D., & Rodriguez, A. (2017). An exploratory investigation of the daily talk time of people with non-fluent aphasia and non-aphasic peers. *International Journal of Speech-Language Pathology, 19*(4), 418–429. https://doi.org/10.1080/17549507.2016.1209558

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics & Phonetics, 30*(7), 489–518. https://doi.org/10.3109/02699206.2016.1145740

Bryant, L., Spencer, E., & Ferguson, A. (2017). Clinical use of linguistic discourse analysis for the assessment of language in aphasia. *Aphasiology, 31*(10), 1105–1126. https://doi.org/10.1080/02687038.2016.1239013

Bzdok, D., Krzywinski, M., & Altman, N. (2017). Machine learning: A primer. *Nature Methods, 14*(12), 1119–1120. https://doi.org/10.1038/nmeth.4526

Cahana-Amitay, D., Albert, M. L., Pyun, S.-B., Westwood, A., Jenkins, T., Wolford, S., & Finley, M. (2011). Language as a stressor in aphasia. *Aphasiology, 25*(5), 593–614. https://doi.org/10.1080/02687038.2010.541469

Casilio, M., Rising, K., Beeson, P. M., Bunton, K., & Wilson, S. M. (2019). Auditory-perceptual rating of connected speech in aphasia. *American Journal of Speech-Language Pathology, 28*(2), 550–568. https://doi.org/10.1044/2018_AJSLP-18-0192

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care — Addressing ethical challenges. *The New England Journal of Medicine, 378*(11), 981–983. https://doi.org/10.1056/NEJMp1714229

***Cho, S., Nevler, N., Ash, S., Shellikeri, S., Irwin, D. J., Massimo, L., Rascovsky, K., Olm, C., Grossman, M., & Liberman, M.** (2021). Automated analysis of lexical features in frontotemporal degeneration. *Cortex, 137,* 215–231. https://doi.org/10.1016/j.cortex.2021.01.012

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

Connaghan, K., Green, J., Paganoni, S., Chan, J., Weber, H., Collins, E., Richburg, B., Eshghi, M., Onnela, J. P., & Berry, J. (2019). Use of Beiwe smartphone app to identify and track speech decline in amyotrophic lateral sclerosis (ALS). *Proceedings of Interspeech 2019,* 4504–4508. https://doi.org/10.21437/Interspeech.2019-3126

***Cordella, C., Dickerson, B. C., Quimby, M., Yunusova, Y., & Green, J. R.** (2017). Slowed articulation rate is a sensitive diagnostic marker for identifying non-fluent primary progressive aphasia. *Aphasiology, 31*(2), 241–260. https://doi.org/10.1080/02687038.2016.1191054

Cordella, C., Gutz, S. E., Eshghi, M., Stipancic, K. L., Schliep, M., Dickerson, B. C., & Green, J. R. (2022). Acoustic and kinematic assessment of motor speech impairment in patients with suspected four-repeat tauopathies. *Journal of Speech, Language, and Hearing Research, 65*(11), 4112–4132. https://doi.org/10.1044/2022_JSLHR-22-00177

***Cordella, C., Quimby, M., Touroutoglou, A., Brickhouse, M., Dickerson, B. C., & Green, J. R.** (2019). Quantification of motor speech impairment and its anatomic basis in primary progressive aphasia. *Neurology, 92*(17), e1992–e2004. https://doi.org/10.1212/WNL.0000000000007367

Duffy, J. R., Hanley, H., Utianski, R., Clark, H., Strand, E., Josephs, K. A., & Whitwell, J. L. (2017). Temporal acoustic measures distinguish primary progressive apraxia of speech from primary progressive aphasia. *Brain and Language, 168,* 84–94. https://doi.org/10.1016/j.bandl.2017.01.012

Duffy, J. R., Strand, E. A., Clark, H., Machulda, M., Whitwell, J. L., & Josephs, K. A. (2015). Primary progressive apraxia of

speech: Clinical features and acoustic and neurologic correlates. *American Journal of Speech-Language Pathology, 24*(2), 88–100. https://doi.org/10.1044/2015_AJSLP-14-0174

Duffy, J. R., Strand, E. A., & Josephs, K. A. (2014). Motor speech disorders associated with primary progressive aphasia. *Aphasiology, 28*(8–9), 1004–1017. https://doi.org/10.1080/02687038.2013.869307

***Faroqi-Shah, Y., Treanor, A., Bernstein Ratner, N., Ficek, B., Webster, K., & Tsapkini, K.** (2020). Using narratives in differential diagnosis of neurodegenerative syndromes. *Journal of Communication Disorders, 85,* Article 105994. https://doi.org/10.1016/j.jcomdis.2020.105994

***Feenaughty, L., Basilakos, A., Bonilha, L., & Fridriksson, J.** (2021). Speech timing changes accompany speech entrainment in aphasia. *Journal of Communication Disorders, 90,* Article 106090. https://doi.org/10.1016/j.jcomdis.2021.106090

Feyereisen, P., Pillon, A., & de Partz, M.-P. (1991). On the measures of fluency in the assessment of spontaneous speech production by aphasic subjects. *Aphasiology, 5*(1), 1–21. https://doi.org/10.1080/02687039108248516

Fraser, K., Rudzicz, F., & Rochon, E. (2013). Using text and acoustic features to diagnose progressive aphasia and its subtypes. In *Proceedings of Interspeech 2013* (pp. 2171–2181). https://doi.org/10.21437/Interspeech.2013-514

***Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., & Rochon, E.** (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex, 55,* 43–60. https://doi.org/10.1016/j.cortex.2012.12.006

Fromm, D., Greenhouse, J., Pudil, M., Shi, Y., & MacWhinney, B. (2022). Enhancing the classification of aphasia: A statistical analysis using connected speech. *Aphasiology, 36*(12), 1492–1519. https://doi.org/10.1080/02687038.2021.1975636

García, A. M., Welch, A. E., Mandelli, M. L., Henry, M. L., Lukic, S., Prioris, M. J. T., Deleon, J., Ratnasiri, B. M., Lorca-Puls, D. L., Miller, B. L., Seeley, W., Vogel, A. P., & Gorno-Tempini, M. L. (2022). Automated detection of speech timing alterations in autopsy-confirmed non-fluent/agrammatic variant primary progressive aphasia. *Neurology, 99*(5), e500–e511. https://doi.org/10.1212/WNL.0000000000200750

***Ghoreishi, Z. S., Azimian, M., Alaghband Rad, J., Khorrami Banaraki, A., Rafiee, S. M., Salavati, M., Shiani, A., & Bayat, N.** (2020). Analysis of connected speech in Persian aphasic patients and its relationship with type and severity of aphasia. *Function and Disability Journal, 15*(3), 141–150. https://doi.org/10.32598/fdj.4.14

Gleichgerrcht, E., Roth, R., Fridriksson, J., den Ouden, D., Delgaizo, J., Stark, B., Hickok, G., Rorden, C., Wilmskoetter, J., Hillis, A., & Bonilha, L. (2021). Neural bases of elements of syntax during speech production in patients with aphasia. *Brain and Language, 222,* Article 105025. https://doi.org/10.1016/j.bandl.2021.105025

Goodglass, H., & Kaplan, E. (1972). *The assessment of aphasia and related disorders.* Lea & Febiger.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *BDAE-3: Boston Diagnostic Aphasia Examination–Third Edition.* Pro-Ed. https://www.proedinc.com/Products/11850/bdae3-boston-diagnostic-aphasia-examinationthird-edition.aspx

Gordon, J. K. (1998). The fluency dimension in aphasia. *Aphasiology, 12*(7–8), 673–688. https://doi.org/10.1080/02687039808249565

Gordon, J. K. (2020). Factor analysis of spontaneous speech in aphasia. *Journal of Speech, Language, and Hearing Research, 63*(12), 4127–4147. https://doi.org/10.1044/2020_JSLHR-20-00340

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

***Gordon, J. K., & Clough, S.** (2022). How do clinicians judge fluency in aphasia? *Journal of Speech, Language, and Hearing Research, 65*(4), 1521–1542. https://doi.org/10.1044/2021_JSLHR-21-00484

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., & Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology, 76*(11), 1006–1014. https://doi.org/10.1212/WNL.0b013e31821103e6

Gosztolya, G., Vincze, V., Tóth, L., Pákáski, M., Kálmán, J., & Hoffmann, I. (2019). Identifying mild cognitive impairment and mild Alzheimer's disease based on spontaneous speech using ASR and linguistic features. *Computer Speech & Language, 53,* 181–197. https://doi.org/10.1016/j.csl.2018.07.007

Grollemund, V., Pradat, P.-F., Querin, G., Delbot, F., Le Chat, G., Pradat-Peyre, J.-F., & Bede, P. (2019). Machine learning in amyotrophic lateral sclerosis: Achievements, pitfalls, and future directions. *Frontiers in Neuroscience, 13,* Article 135. https://doi.org/10.3389/fnins.2019.00135

Grossman, M. (2012). The non-fluent/agrammatic variant of primary progressive aphasia. *The Lancet Neurology, 11*(6), 545–555. https://doi.org/10.1016/S1474-4422(12)70099-6

Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2017a). Triangulation of language-cognitive impairments, naming errors and their neural bases post-stroke. *NeuroImage: Clinical, 17,* 465–473. https://doi.org/10.1016/j.nicl.2017.10.037

Halai, A. D., Woollams, A. M., & Lambon Ralph, M. A. (2017b). Using principal component analysis to capture individual differences within a unified neuropsychological model of chronic post-stroke aphasia: Revealing the unique neural correlates of speech fluency, phonology and semantics. *Cortex, 86,* 275–289. https://doi.org/10.1016/j.cortex.2016.04.016

***Haley, K. L., Jacks, A., Jarrett, J., Taylor, R., Cunningham, K. T., Gorno-Tempini, M. L., & Henry, M. L.** (2021). Speech metrics and samples that differentiate between nonfluent/agrammatic and logopenic variants of primary progressive aphasia. *Journal of Speech, Language, Hearing Research, 64*(3), 754–775. https://doi.org/10.1044/2020_JSLHR-20-00445

***Hardy, C., Buckley, A., Downey, L., Lehmann, M., Zimmerer, V., Varley, R., Crutch, S., Rohrer, J., Warrington, E., & Warren, J.** (2016). The language profile of behavioral variant frontotemporal dementia. *Journal of Alzheimer's Disease, 50*(2), 359–371. https://doi.org/10.3233/JAD-150806

***Harmon, T. G., Jacks, A., & Haley, K. L.** (2019). Speech fluency in acquired apraxia of speech during narrative discourse: Group comparisons and dual-task effects. *American Journal of Speech-Language Pathology, 28*(2S), 905–914. https://doi.org/10.1044/2018_AJSLP-MSC18-18-0107

Harmon, T. G., Jacks, A., Haley, K. L., & Faldowski, R. A. (2016). Listener perceptions of simulated fluent speech in nonfluent aphasia. *Aphasiology, 30*(8), 922–942. https://doi.org/10.1080/02687038.2015.1077925

Henry, M. L., & Gorno-Tempini, M. L. (2010). The logopenic variant of primary progressive aphasia. *Current Opinion in Neurology, 23*(6), 633–637. https://doi.org/10.1097/WCO.0b013e32833fb93e

Ingram, R. U., Halai, A. D., Pobric, G., Sajjadi, S., Patterson, K., & Lambon Ralph, M. A. (2020). Graded, multidimensional intra- and intergroup variations in primary progressive aphasia and post-stroke aphasia. *Brain, 143*(10), 3121–3135. https://doi.org/10.1093/brain/awaa245

Jacks, A., Haley, K. L., Bishop, G., & Harmon, T. G. (2019). Automated speech recognition in adult stroke survivors: Comparing human and computer transcriptions. *Folia Phoniatrica et Logopaedica, 71*(5–6), 286–296. https://doi.org/10.1159/000499156

Josephs, K. A., Duffy, J. R., Martin, P. R., Stephens, Y. C., Atulkumar Singh, N., Clark, H. M., Botha, H., Lowe, V. J., Whitwell, J. L., & Utianski, R. L. (2023). Acoustic analysis and neuroimaging correlates of diadochokinetic rates in mild–moderate primary progressive apraxia of speech. *Brain and Language, 240,* Article 105254. https://doi.org/10.1016/j.bandl.2023.105254

Josephs, K. A., Duffy, J. R., Strand, E. A., Machulda, M. M., Senjem, M. L., Gunter, J. L., Schwarz, C. G., Reid, R. I., Spychalla, A. J., Lowe, V. J., Jack, C. R., Jr., & Whitwell, J. L. (2014). The evolution of primary progressive apraxia of speech. *Brain, 137*(10), 2783–2795. https://doi.org/10.1093/brain/awu223

Josephs, K. A., Duffy, J. R., Strand, E. A., Machulda, M. M., Senjem, M. L., Master, A. V., Lowe, V. J., Jack, C. R., Jr., & Whitwell, J. L. (2012). Characterizing a neurodegenerative syndrome: Primary progressive apraxia of speech. *Brain, 135*(5), 1522–1536. https://doi.org/10.1093/brain/aws032

Kertesz, A. (2007). *Western Aphasia Battery–Revised*. PsychCorp.

***Kim, H., Kintz, S., & Wright, H. H.** (2021). Development of a measure of function word use in narrative discourse: Core lexicon analysis in aphasia. *International Journal of Language & Communication Disorders, 56*(1), 6–19. https://doi.org/10.1111/1460-6984.12567

***Kim, H., Kintz, S., Zelnosky, K., & Wright, H. H.** (2019). Measuring word retrieval in narrative discourse: Core lexicon in aphasia. *International Journal of Language & Communication Disorders, 54*(1), 62–78. https://doi.org/10.1111/1460-6984.12432

Kong, A. P.-H., & Wong, C. W.-Y. (2018). An integrative analysis of spontaneous storytelling discourse in aphasia: Relationship with listeners' rating and prediction of severity and fluency status of aphasia. *American Journal of Speech-Language Pathology, 27*(4), 1491–1505. https://doi.org/10.1044/2018_AJSLP-18-0015

***Kong, A. P.-H., Whiteside, J., & Bargmann, P.** (2016). The Main Concept Analysis: Validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. *Logopedics Phoniatrics Vocology, 41*(3), 129–141. https://doi.org/10.3109/14015439.2015.1041551

Kristinsson, S., Zhang, W., Rorden, C., Newman-Norlund, R., Basilakos, A., Bonilha, L., Yourganov, G., Xiao, F., Hillis, A., & Fridriksson, J. (2021). Machine learning-based multimodal prediction of language outcomes in chronic aphasia. *Human Brain Mapping, 42*(6), 1682–1698. https://doi.org/10.1002/hbm.25321

Laures-Gore, J. S., & DeFife, L. C. (2013). Perceived stress and depression in left and right hemisphere post-stroke patients. *Neuropsychological Rehabilitation, 23*(6), 783–797. https://doi.org/10.1080/09602011.2013.811087

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

Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology, 5*(1), 96–116. https://doi.org/10.1002/lio2.354

***Mack, J., Chandler, S., Meltzer-Asscher, A., Rogalski, E., Weintraub, S., Mesulam, M., & Thompson, C.** (2015). What do pauses in narrative production reveal about the nature of word retrieval deficits in PPA? *Neuropyschologia, 77,* 211–222. https://doi.org/10.1016/j.neuropsychologia.2015.08.019

MacWhinney, B. (2000). *Tools for analyzing talk part 1: The CHAT transcription format*. Lawrence Erlbaum Associates.

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

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

MacWhinney, B., & Wagner, J. (2010). Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprachsforschung: Online-Zeitschrift Zur Verbalen Interaktion, 11,* 154–173.

***Manning, M., & Franklin, S.** (2016). Cognitive grammar and aphasic discourse. *Clinical Linguistics & Phonetics, 30*(6), 417–432. https://doi.org/10.3109/02699206.2015.1128981

***Marcotte, K, Graham, N. L., Fraser, K. C., Meltzer, J. A., Tang-Wai, D. F., Chow, T. W., Freedman, M., Leonard, C., Black, S. E., & Rochon, E.** (2017). White matter disruption and connected speech in non-fluent and semantic variants of primary progressive aphasia. *Dementia and Geriatric Cognitive Disorders Extra, 7*(1), 52–73. https://doi.org/10.1159/000456710

***Martinez-Ferreiro, S., González, E., Clari, V., & Bastiaanse, R.** (2017). Quantifying connected discourse in Spanish-speaking individuals with aphasia: The case of mixed aphasias. *Journal of Neurolinguistics, 44,* 38–53. https://doi.org/10.1016/j.jneuroling.2017.03.001

***Matias-Guiu, J. A., Suárez-Coalla, P., Yus, M., Pytel, V., Hernández-Lorenzo, L., Delgado-Alonso, C., Delgado-Álvarez, A., Gómez-Ruiz, N., Polidura, C., Cabrera-Martín, M. N., Matías-Guiu, J., & Cuetos, F.** (2022). Identification of the main components of spontaneous speech in primary progressive aphasia and their neural underpinnings using multimodal MRI and FDG-PET imaging. *Cortex, 146,* 141–160. https://doi.org/10.1016/j.cortex.2021.10.010

Mesulam, M.-M. (2001). Primary progressive aphasia. *Annals of Neurology, 49*(4), 425–432. https://doi.org/10.1002/ana.91

Mesulam, M.-M., Wieneke, C., Rogalski, E., Cobia, D., Thompson, C., & Weintraub, S. (2009). Quantitative template for subtyping primary progressive aphasia. *Archives of Neurology, 66*(12), 1545–1551. https://doi.org/10.1001/archneurol.2009.288

Mesulam, M.-M., Wieneke, C., Thompson, C., Rogalski, E., & Weintraub, S. (2012). Quantitative classification of primary progressive aphasia at early and mild impairment stages. *Brain, 135*(5), 1537–1553. https://doi.org/10.1093/brain/aws080

Miller, J., Andriacchi, K., & Nockerts, A. (2015). *Assessing language production using SALT software: A clinician's guide to language sample analysis* (*2nd ed.*) [Computer software].

Miller, M. I., Shih, L. C., & Kolachalama, V. B. (2023). Machine learning in clinical trials: A primer with applications to neurology. *Neurotherapeutics, 20*(4), 1066–1080. https://doi.org/10.1007/s13311-023-01384-2

***Mirman, D., Kraft, A., Harvey, D., Brecher, A., & Schwartz, M.** (2019). Mapping articulatory and grammatical subcomponents of fluency deficits in post-stroke aphasia. *Cognitive, Affective, & Behavioral Neuroscience, 19*(5), 1286–1298. https://doi.org/10.3758/s13415-019-00729-9

Montembeault, M., Brambati, S. M., Gorno-Tempini, M. L., & Migliaccio, R. (2018). Clinical, anatomical, and pathological features in the three variants of primary progressive aphasia: A review. *Frontiers in Neurology, 9,* Article 692. https://doi.org/10.3389/fneur.2018.00692

National Institute on Deafness and Other Communication Disorders. (2017). *National Institute on Deafness and Other Communication Disorders (NIDCD) fact sheet: Aphasia*. https://www.nidcd.nih.gov/sites/default/files/Documents/health/voice/Aphasia.pdf [PDF]

Nevler, N., Ash, S., Irwin, D., Liberman, M., & Grossman, M. (2019a). P2-254: Automated prosodic markers in natural connected speech of ALS and ALS-FTD are associated with cognitive impairment. *Alzheimer's and Dementia, 15*(7S, Part 13), P681–P682. https://doi.org/10.1016/j.jalz.2019.06.2661

***Nevler, N., Ash, S., Irwin, D. J., Liberman, M., & Grossman, M.** (2019b). Validated automatic speech biomarkers in primary progressive aphasia. *Annals of Clinical Translational Neurology, 6*(1), 4–14. https://doi.org/10.1002/acn3.653

Nevler, N., Ash, S., McMillan, C., Elman, L., McCluskey, L., Irwin, D. J., Cho, S., Liberman, M., & Grossman, M. (2020). Automated analysis of natural speech in amyotrophic lateral sclerosis spectrum disorders. *Neurology, 95*(12), e1629–e1639. https://doi.org/10.1212/WNL.0000000000010366

Nozari, N., & Faroqi-Shah, Y. (2017). Investigating the origin of nonfluency in aphasia: A path modeling approach to neuropsychology. *Cortex, 95,* 119–135. https://doi.org/10.1016/j.cortex.2017.08.003

Ogar, J. M., Dronkers, N. F., Brambati, S. M., Miller, B. L., & Gorno-Tempini, M. L. (2007). Progressive nonfluent aphasia and its characteristic motor speech deficits. *Alzheimer Disease & Associated Disorders, 21*(4), S23–S30. https://doi.org/10.1097/WAD.0b013e31815d19fe

Parjane, N., Cho, S., Ash, S., Cousins, K. A. Q., Shellikeri, S., Liberman, M., Shaw, L. M., Irwin, D. J., Grossman, M., & Nevler, N. (2021). Digital speech analysis in progressive supranuclear palsy and corticobasal syndromes. *Journal of Alzheimer's Disease, 82*(1), 33–45. https://doi.org/10.3233/JAD-201132

Park, H., Rogalski, Y., Rodriguez, A. D., Zlatar, Z., Benjamin, M., Harnish, S., Bennett, J., Rosenbek, J. C., Crosson, B., & Reilly, J. (2011). Perceptual cues used by listeners to discriminate fluent from nonfluent narrative discourse. *Aphasiology, 25*(9), 998–1015. https://doi.org/10.1080/02687038.2011.570770

Poeck, K. (1989). Fluency. In C. Code (Ed.), *The characteristics of aphasia* (pp. 23–32). Taylor & Francis.

Poole, M. L., Brodtmann, A., Darby, D., & Vogel, A. P. (2017). Motor speech phenotypes of frontotemporal dementia, primary progressive aphasia, and progressive apraxia of speech. *Journal of Speech, Language, and Hearing Research, 60*(4), 897–911. https://doi.org/10.1044/2016_JSLHR-S-16-0140

Prins, R., & Bastiaanse, R. (2004). Review. *Aphasiology, 18*(12), 1075–1091. https://doi.org/10.1080/02687030444000534

Reeves, M. J., Bushnell, C. D., Howard, G., Gargano, J. W., Duncan, P. W., Lynch, G., Khatiwoda, A., & Lisabeth, L. (2008). Sex differences in stroke: Epidemiology, clinical presentation, medical care, and outcomes. *Lancet Neurology, 7*(10), 915–926. https://doi.org/10.1016/S1474-4422(08)70193-5

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. https://doi.org/10.1006/brln.1999.2285

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37*(3), 440–479. https://doi.org/10.1016/0093-934x(89)90030-8

***Sajjadi, S. A., Patterson, K., Tomek, M., & Nestor, P. J.** (2012). Abnormalities of connected speech in the non-semantic variants of primary progressive aphasia. *Aphasiology, 26*(10), 1219–1237. https://doi.org/10.1080/02687038.2012.710318

Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making, 7*(1), Article 16. https://doi.org/10.1186/1472-6947-7-16

***Sitek, E. J., Kluj-Kozłowska, K., Barczak, A., Kozłowski, M., Wieczorek, D., Przewłócka, A., Narozańska, E., Dąbrowska, M., Barcikowska, M., & Sławek, J.** (2015). Overlapping and distinguishing features of descriptive speech in Richardson variant of progressive supranuclear palsy and non-fluent progressive aphasia. *Postepy Psychiatrii i Neurologii, 24*(2), 62–67. https://doi.org/10.1016/j.pin.2015.06.003

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

Stark, B. C., Dutta, M., Murray, L. L., Fromm, D., Bryant, L., Harmon, T. G., Ramage, A. E., & Roberts, A. C. (2021). Spoken discourse assessment and analysis in aphasia: An international survey of current practices. *Journal of Speech, Language, and Hearing Research, 64*(11), 4366–4389. https://doi.org/10.1044/2021_JSLHR-20-00708

Stegmann, G. M., Hahn, S., Liss, J., Shefner, J., Rutkove, S., Shelton, K., Duncan, C. J., & Berisha, V. (2020). Early detection and tracking of bulbar changes in ALS via frequent and remote speech analysis. *npj Digital Medicine, 3*(1), Article 132. https://doi.org/10.1038/s41746-020-00335-x

Strand, E. A., Duffy, J. R., Clark, H. M., & Josephs, K. A. (2014). The apraxia of speech rating scale: A tool for diagnosis and description of apraxia of speech. *Journal of Communication Disorders, 51,* 43–50. https://doi.org/10.1016/j.jcomdis.2014.06.008

Tavabi, N., Stück, D., Signorini, A., Karjadi, C., Al Hanai, T., Sandoval, M., Lemke, C., Glass, J., Hardy, S., Lavallee, M., Wasserman, B., Ang, T. F. A., Nowak, C. M., Kainkaryam, R., Foschini, L., & Au, R. (2022). Cognitive digital biomarkers from automated transcription of spoken language. *The Journal of Prevention of Alzheimer's Disease, 9*(4), 791–800. https://doi.org/10.14283/jpad.2022.66

Tee, B. L., & Gorno-Tempini, M. L. (2019). Primary progressive aphasia: A model for neurodegenerative disease. *Current Opinion in Neurology, 32*(2), 255–265. https://doi.org/10.1097/WCO.0000000000000673

Themistocleous, C., Ficek, B., Webster, K., den Ouden, D.-B., Hillis, A. E., & Tsapkini, K. (2021). Automatic subtyping of individuals with primary progressive aphasia. *Journal of Alzheimer's Disease, 79*(3), 1185–1194. https://doi.org/10.3233/JAD-201101

***Themistocleous, C., Webster, K., Afthinos, A., & Tsapkini, K.** (2021). Part of speech production in patients with primary progressive aphasia: An analysis based on natural language processing. *American Journal of Speech-Language Pathology, 30*(1S), 466–480. https://doi.org/10.1044/2020_AJSLP-19-00114

Thompson, C. K., Ballard, K. J., Tait, M. E., Weintraub, S., & Mesulam, M. (1997). Patterns of language decline in nonfluent primary progressive aphasia. *Aphasiology, 11*(4–5), 297–321. https://doi.org/10.1080/02687039708248473

***Thompson, C. K., Cho, S., Hsu, C.-J., Wieneke, C., Rademaker, A., Weitner, B. B., Mesulam, M. M., & Weintraub, S.** (2012).

Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology, 26*(1), 20–43. https://doi.org/10.1080/02687038.2011.584691

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language, 162,* 60–71. https://doi.org/10.1016/j.bandl.2016.08.004

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., . . . Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

Trupe, E. (1984). Reliability of rating spontaneous speech in the Western Aphasia Battery: Implications for classification. *Clinical Aphasiology: Proceedings of the Conference 1984,* 55–69. http://aphasiology.pitt.edu/id/eprint/796

Vandenberghe, R. (2016). Classification of the primary progressive aphasias: Principles and review of progress since 2011. *Alzheimer's Research & Therapy, 8*(1), Article 16. https://doi.org/10.1186/s13195-016-0185-y

Veritas Health Innovation. (2022). *Covidence systematic review software* [Computer software]. http://www.covidence.org

Vermeulen, J., Bastiaanse, R., & Van Wageningen, B. (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language, 36*(2), 252–274. https://doi.org/10.1016/0093-934x(89)90064-3

Wagenaar, E., Snow, C., & Prins, R. (1975). Spontaneous speech of aphasic patients: A psycholinguistic analysis. *Brain and Language, 2*(3), 281–303. https://doi.org/10.1016/s0093-934x(75)80071-x

Wilson, S. M., Entrup, J. L., Schneck, S. M., Onuscheck, C. F., Levy, D. F., Rahman, M., Willey, E., Casilio, M., Yen, M., Brito, A. C., Kam, W., Davis, L. T., de Riesthal, M., & Kirshner, H. S. (2022). Recovery from aphasia in the first year after stroke. *Brain, 146*(3), 1021–1039. https://doi.org/10.1093/brain/awac129

Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., Miller, B. L., & Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain, 133*(7), 2069–2088. https://doi.org/10.1093/brain/awq129

Xue, C., Karjadi, C., Paschalidis, I. Ch., Au, R., & Kolachalama, V. B. (2021). Detection of dementia on voice recordings using deep learning: A Framingham Heart Study. *Alzheimer's Research & Therapy, 13*(1), Article 146. https://doi.org/10.1186/s13195-021-00888-3

Yoon, C. H., Torrance, R., & Scheinerman, N. (2022). Machine learning in medicine: Should the pursuit of enhanced interpretability be abandoned? *Journal of Medical Ethics, 48*(9), 581–585. https://doi.org/10.1136/medethics-2020-107102

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders, 45*(1), 27–36. https://doi.org/10.1044/jshd.4501.27

Yunusova, Y., Graham, N. L., Shellikeri, S., Phuong, K., Kulkarni, M., Rochon, E., Tang-Wai, D. F., Chow, T. W., Black, S. E., Zinman, L. H., & Green, J. R. (2016). Profiling speech and pausing in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). *PLOS ONE, 11*(1), Article e0147573. https://doi.org/10.1371/journal.pone.0147573

***Zhang, M., Geng, L., Yang, Y., & Ding, H.** (2021). Cohesion in the discourse of people with post-stroke aphasia. *Clinical Linguistics & Phonetics, 35*(1), 2–18. https://doi.org/10.1080/02699206.2020.1734864

Zhao, Y., Halai, A., & Ralph, M. (2020). Evaluating the granularity and statistical structure of lesions and behaviour in post-stroke aphasia. *Brain Communications, 2*(2). https://doi.org/10.1093/braincomms/fcaa062

***Zimmerer, V. C., Hardy, C. J. D., Eastman, J., Dutta, S., Varnet, L., Bond, R. L., Russell, L., Rohrer, J. D., Warren, J. D., & Varley, R. A.** (2020). Automated profiling of spontaneous speech in primary progressive aphasia and behavioral-variant frontotemporal dementia: An approach based on usage-frequency. *Cortex, 133,* 103–119. https://doi.org/10.1016/j.cortex.2020.08.027

***Zimmerer, V. C., Newman, L., Thomson, R., Coleman, M., & Varley, R.** (2018). Automated analysis of language production in aphasia and right-hemisphere damage: Frequency and collocation strength. *Aphasiology, 32*(11), 1267–1283. https://doi.org/10.1080/02687038.2018.1497138

Zraick, R. I., & Boone, D. R. (1991). Spouse attitudes toward the person with aphasia. *Journal of Speech and Hearing Research, 34*(1), 123–128. https://doi.org/10.1044/jshr.3401.123